Proceedings of the 19th International Conference on

Flexible Automation and
Intelligent Manufacturing

FAIM 2009

July 6th – July 8th, 2009
Teesside University, Middlesbrough, UK

Editor: Farhad Nabhani
Co-editors: Catherine Frost
           Sara Zarei
           Munir Ahmad
           William. G. Sullivan
PREFACE

The 19th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM), has a rich and time-tested tradition of excellence culminating in the FAIM2009 conference at Teesside University in Middlesbrough, England. The 1st International FAIM conference in 1991 resulted from collaboration between the University in Limerick (UL) and the Virginia Polytechnic Institute. Prof. Munir Ahmad (then UL), and Prof. Bill Sullivan (then Virginia Tech.) have been central to the development of the Conference since its inception. Since 1991 FAIM has been hosted in many prestigious Universities on both sides of the Atlantic:

1991 University of Limerick, Limerick, Ireland
1992 Virginia Tech (in Washington DC), USA
1993 University of Limerick, Limerick, Ireland
1994 Virginia Tech, Blacksburg, Virginia, USA
1995 University of Stuttgart, Germany
1996 Georgia Institute of Technology, Atlanta, USA
1997 University of Teesside, Middlesbrough, UK
1998 Portland State University, Portland, Oregon, USA
1999 Tilburg University, Tilburg, Netherlands
2000 University of Maryland, Maryland, USA
2001 Dublin City University, Dublin, Ireland
2002 Dresden University of Technology, Dresden, Germany
2003 University of South Florida, Tampa, USA
2004 Ryerson University, Toronto, Canada
2005 University of Duesto, Bilbao, Spain
2006 University of Limerick, Limerick, Ireland
2007 Penn State Great Valley, Philadelphia, USA
2008 University of Skövde, Sweden

This year, the FAIM conference is organized by the Teesside University, Middlesbrough. Papers were accepted from authors from 37 countries. All accepted 192 papers for this year’s conference proceedings were rigorously peer reviewed and carefully selected from more than 275 abstracts submitted to the Conference.

The topics covered in the papers include:

- Bio-Engineering
- Business Process Re-Engineering
- Business Process Re-Engineering, Lean and Agile Manufacture
- CAD/CAM Flexible Manufacturing, CIM
- Change Management
- E-manufacturing
- Engineering for Sustainability
- Environmental Engineering
- Global Manufacturing
The FAIM 2009 conference focuses on “lean manufacturing and services” covering current research, best practices and future trends within the areas of global competitiveness and rapidly advancing technologies in flexible automation, information management, and intelligent manufacturing. FAIM 2009 provides a forum for both researchers and practitioners in the field of Flexible Automation and Intelligent Manufacturing to share the latest research, developments, and practices in the areas of the conference scope. Papers included in the proceedings address applications of these technologies and methods to the manufacturing environment, from industrial and academic perspectives.

Manufacturing today must adjust to dynamic markets which demand excellence in areas of cost, quality, throughput, innovativeness, and time to market. Manufacturing companies must master emerging and rapidly advancing technologies in order to remain competitive. They are challenged to achieve short product development cycles, expand product lines, and combat shrinking product life cycles in a global economy marked by rapid fluctuations and increasing competitions. This tradition will be continued next year in San Francisco, USA where they will host the next FAIM conference.
ACKNOWLEDGEMENTS

As the chairperson of the 19th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2009), I am indebted to many individuals for their contribution, support, and endorsement. I wish to thank all keynote speakers who shared their views and visions towards flexible automation and intelligent manufacturing. I also wish to acknowledge, with many thanks, the contributions of all authors who presented their work at the conference and submitted quality papers for this proceedings publication. My special thanks are extended to all FAIM 2009 peer reviewers and session moderators who helped ensure the high quality of the conference.

I would further like to acknowledge the contribution of the International Program and Local FAIM Organizing Committees, for their work in scheduling, the preparation of the papers, session moderation, organization of the social events, and the general operation involved in running this international conference. I would like to thank Caterpillar Ltd and Cummins Engine UK for hosting the FAIM industrial study visit.

The continued success of FAIM can be attributed to the guidance of its two founders, Prof. M. M. Ahmad and Prof. W. G. Sullivan, whom I would like to acknowledge and thank. I would also like to thank many former organisers of FAIM for their heart warming advice and support.

I would like to acknowledge the role of Teesside University and wish to express my sincere thanks to Professor Graham Henderson, the Vice Chancellor and Professor Simon Hodgson, Dean of School of Science and Technology for support in running such a prestigious conference, and thank you to all the support staff who often go unnoticed.

Although there are many who deserve credit for their various contributions to the success of FAIM 2009, there is someone who deserves special recognition and big thanks for her outstanding efforts, tireless work and commitment, Ms. Catherine Frost the Conference Organizer. Without her hard work, this conference and its proceedings would have been impossible.

Also I wish to acknowledge the excellent editorial team Sara Zarei, Martin McKie and Dr Thanos Klonis for their work as the local internal conference support team; their patience and enthusiasm for this editing this work are sincerely appreciated.

Professor Farhad Nabhani
Chair of FAIM 2009 Conference
FAIM 2009 TEAM AND COMMITTEES

CHAIR

- Prof Farhad Nabhani, University of Teesside, UK

CO-CHAIRS

- Prof. Munir Ahmad, University of Teesside, Middlesbrough, U.K
- Prof. Zulf Ali, University of Teesside, Middlesbrough, U.K
- Prof Simon Hodgson, University of Teesside, Middlesbrough, UK
- Prof. Bill Sullivan, Virginia Polytechnic Institute and State University, U.S.A.

ORGANISATION COMMITTEE

- Elaine Bradwell
- Simon Hodgson
- Jayne Laverick
- Paul McGrath
- Farhad Nabhani
- Mark White
- Amanda Wright

SCIENTIFIC COMMITTEE

- Esther Alvarez, Universidad de Deusto, Bilbao, Spain
- Jalal Ashayeri, Tilburg University, The Netherlands
- Farid Benyahia, College of Engineering, Qatar University
- Shahroz Connolly, University of Teesside, UK
- Marjan Darabi, Damyaran Arak Ltd, Iran
- Leo J De Vin, University of Skovde, Sweden
- S M T Fatemi Ghomi, Amirkabir University of Technology, Iran
- Farnaz Ganjeizadeh, California State University, USA
- Bill Gaughran, University of Limerick, Ireland
- Mark Gerrard, University of Teesside, UK
- Sunderesh S Heragu, University of Louisville, USA
- George Q Huang, University of Hong Kong, HK
- Winifred Ijomah, University of Strathclyde, UK
- Riham Khalil, De Montfort University, UK
- Ravi Y Kumar, National Institute of Technology, Warangal, India
- Feng Lin, Tsinghua, P.R.China
- Richard Lindeke, University of Minnesota Duluth, USA
- Gonzales Mejia, University De Los Andes, Colombia
- Norhamidi Muhamad, University of Kebangasaan, Malaysia
- Stephen T. Newman, Bath University, UK
• Chike F Oduoza, University of Wolverhampton, UK
• David L Olson, University of Nebraska, USA
• C S P Rao, National Institute of Technology, Warangal, India
• Olli-Pekka Hilmola, Lappeenranta University of Technology, Finland
• David Russell, Penn State Great Valley, Philadelphia, USA
• Michael A Saliba, University of Malta, Malta
• Ignacio Eguia Salinas, Universidad de Sevilla, Spain
• Thomson Sarkodie-Gyan, University of Texas, El Paso, USA
• Mohammed Sarwar, Northumbria University, UK
• Paul Shelton, University of Teesside, UK
• Piotr Skawinski, Warsaw University of Technology, Poland
• Dusan Sormaz, Ohio University, USA
• Janis P Terpenny, Virginia Tech, USA
• Uday Venkatadri, Dalhousie University, Canada
• Lihui Wang, National Research Council, Canada
• Gerald Weigert, Technische Universitat Dresden, Germany
• David A Wyrick, Texas Tech University, USA

EDITORIAL COMMITTEE

• Athanosis Klonis
• Martin Mckie
• Sara Zarei

CONFERENCE SECRETARY

• Catherine Frost
# Table of Contents

## Keynote Lecturers

**Engineering for Sustainability**  
Graham Hillier ................................................................. 1

**Successful Continuous Improvement – How to make and sustain improvements**  
Keith Copeland ............................................................. 2

**RFID-Enabled Wireless Manufacturing Execution Platform**  
George Q. Huang .......................................................... 3

**Process Industry in the UK and in a regional context**  
Stan Higgins ................................................................. 4

**Where lies the Factory of the Future?**  
David W Russell ............................................................. 5

## Bio. Engineering

**Framework of Grinding Process Modeling and Simulation Based on Microscopic Interaction Analysis**  
Xuekun Li, Yiming Rong .................................................. 6

**Clean Technology approach for use of micro-bioreactor in Bioprocess Engineering**  
Pattanathu K.S.M. Rahman, Godfrey Pasirayi, Vincent Auger, and Zulf Ali .................. 14

**Contact Area, Pressure Distribution, and Mechanical Stability in External Arthrodesis of the Ankle Joint**  
C.J. Connor, F. Nabhani and S. Hodgson ........................... 23

**Bracing Shoulder Injuries caused by Rear impacts.**  
D J Hughes, F Nabhani and S Hodgson ............................... 29

## Business Process Re-Engineering

**Inventory System Improvement in ERP Implementation Projects**  
Michael Boyce, Dr. Nasreddin Dhafr & Robert Lee ..................... 37

**Embedding ERP in an Engineering Projects Environment**  
Robert Lee, Michael Boyce, Ahmed Abbas and Nasreddin Dhafr ................. 45

**Design of a Generic Methodology Framework for Process Diagnostics Management**  
Jiri Tupa and Chike Oduoza .............................................. 52

**Global Manufacturing: Creating the Balance between Local and Global Markets**  
Parisa Gilani and Yasamin Razeghi .................................... 60
Potential of Benchmarking
Munir Ahmad ............................................................................................................................. 67

**BUSINESS PROCESS RE-ENGINEERING, LEAN & AGILE MANUFACTURE**

Moving towards Creating an Organic Lean Manufacturing Organisation
Sadri Gilani and Yasamin Razeghi ............................................................................................................. 75

The strategic importance of ERP screening process: Evidence from the forging industry
Athanosis Klonis and Ruben Pinedo-Cuenca ......................................................................................... 81

Simultaneous QFD-Based Optimization of Manufacturing System for Multiproduct Operation
George M. Bacioiu and Zbigniew J. Pasek .......................................................................................... 89

**CAD/CAM FLEXIBLE MANUFACTURING, CIM**

CAD of Main Spindle and Feed Drives for CNC Machine Tools
Zoran Pandilov and Vladimir Dukovski .................................................................................................. 97

Interaction of natural inspired methods in process planning
Cornelia Teich, Daniel Kretz, Matthias Richter, Joerg Militzer and Tim Neumann ............................ 105

Computerized System for Progressive Die Layout Design
Mohamad Ali Farsi and B. Arezoo .......................................................................................................... 112

Implementation of a Modular Supervision System for a Flexible Manufacturing Cell
Joao C.E. Ferreira, Hugo G. Santos, Marcelo T. dos Santos and Léo Schirmer ................................. 122

Cooperative robot kinematics to perform load sharing tasks
Cristiane P. Tonetto and Altamir Dias ................................................................................................. 130

On parametric toolpath design of a robot serving a press-brake
Nikolaos Kontolatis, George-Christopher Vosniakos and Konstantinos Kyriakopoulos .................... 139

The Digital Factory Concept from Manufacturing to Services Environment
Fabio Fruggiero and Alfredo Lambiase ............................................................................................... 147

Object techniques in an integration of technical and organizational production preparation processes
Cezary Grabowik ........................................................................................................................................ 155

A STEP-NC Process Planning System for CNC Turning Operations
X. Zhang, R. Liu, A. Nassehi and S. T. Newman ................................................................................ 163

Mathematical model of 6-axis CNC spiral bevel and hypoid milling machine
Piotr Skawinski ........................................................................................................................................ 171
CHANGE MANAGEMENT

Corporate Social Responsibility – a new trend in company management
Marcela Srchová, Lilia Dvořáková and Tereza Kadlecová ...................................................... 179

E-MANUFACTURING

Web-DPP: A Distributed Process Planning System for Adaptive Machining Operations
Lihui Wang, Ji Ma and Hsi-Yung Feng ................................................................. 186

Implementation of Automobile Spare Tire Inserting System using Augmented Reality
Hong-Seok Park, Jin-Woo Park and Hung-Won Choi ................................................................. 194

Design of Collaborative BOM Management System for Small and Medium Enterprises
Bo Hyun Kim, So Young Jung, Jae Yong Baek, Seok Woo Lee, and Hun Zong Choi .......... 202

ENGINEERING FOR SUSTAINABILITY

Active Disassembly application to enhance Design-for-Remanufacturing
J D Chiodo and W L Ijomah ................................................................................... 210

The Fault-Ride-Through Capacity of Doubly-Fed Induction Generators (DFIGs) for Wind Applications
Wenping Cao and Jessada Stevenson ........................................................................ 218

Predictive Simulation for Model Based Energy Consumption Optimisation in Manufacturing System and Machine Control
Anton Dietmair, Alexander Verl, and Philipp Eberspaecher ................................................... 226

Bionics – An Inspiration for Intelligent Manufacturing and Engineering
Bill Gaughran and Sonya Quinn ...................................................................................... 234

Environmental Management Systems in Smaller Enterprises (Problems and Solutions)
Mike Gallagher, William Gaughran and Pat Phealan ......................................................... 242

ENVIRONMENTAL ENGINEERING

Using biofiltration to clean waste gases
A M Gerrard, J Skalicky and J Paca .............................................................................. 250

Voluntary Environmental Instruments - their benefits and usage in the Czech Republic
Tereza Kadlecová (nee Fidlerová), Lilia Dvořáková and Marcela Srchová ......................... 257

GLOBAL MANUFACTURING

Manufacturing of a Pick-and-Place Robot with Inverse Dynamic and Robust Controllers
S. Saeed Mirian, S. Mohsen Safavi ................................................................. 265
HEALTH AND SAFETY

A Survey on Impacts of Shift Work to Shop Floor Workers’ Safety and Health in a Manufacturing Company
Baba Md Deros, Nor Kamaliana Khamis, Ahmad Rasdan Ismail, and Asri Ludin ................274

Health and safety perspective on medical waste management in a developing country: A case study of Dhaka city
M. A. Patwary, W. T. O’Hare, G. Street, K. M. Elahi, S. S. Hossain and M. H. Sarker ..........282

High Efficiency Cloning of Drug Resistance Genes Via Safe Vector Host Systems
Connolly S.S. and Connolly S.N. ............................................................................................ 291

HUMAN FACTORS AND HUMAN RESOURCES IN MANUFACTURING

Introducing a lean recruitment framework for manufacturing
Paul M Gibbons ..........................................................295

The Demographic Change in Firms of Western Industry Nations - Managerial Implications for Working Systems Design
Marion A, Weissenberger-Eibl and Johann Schwenk ................................................303

Analysis of the Methods Time Measurement (MTM) Methodology through its Application in Manufacturing Companies
Denis L. M. de Almeida and João C. E. Ferreira .................................................................311

A Neural Network Approach for Measuring the Impact of Knowledge Management on Organizational Performance of Turkish SMEs
Ali Turkyilmaz, Erkan Bayraktar, Halil Zaim and Selim Zaim .......................................319

INDUSTRIAL AUTOMATION AND PROCESS CONTROL

A Metric for Rating the Effectiveness of Installed Industrial Automation Systems using an Adjusted Function Point Analysis
David W. Russell .................................................................................................................... 328

CAN Protocol in diffuse controlling system based on B&R hardware
J. Swider, G. Kost, G. Wszelek and D. Reclik ...............................................................336

Fruit Color and Shape Grading Using a New Vision System
H. Golnabi, R. Jafari, A. Haghighatzadeh and E. Jali .........................................................343

Feasibility of an AIDC System within a Manufacturing SME
Sara Zarei, Simon Hodgson and Farhad Nabhani .............................................................351

Design of a Cobot with Three Omni-Wheels
Z. M. Bi and Lihui Wang ........................................................................................................360
Surface Monitoring by Color Reflectance Measurements
H. Golnabi, A. Haghighehtadeh and S. Dabagh ................................................................. 368

Semi-Open Queuing Network Approach for Modeling Autonomous Vehicle Storage and
Retrieval System
Banu Y. Ekren, Sunderesh S. Heragu, Ananth Krishnamurthy and Charles J. Malmborg ...... 376

A Method for the Calibration of a 3-D Laser Scanner
Cesare Rossi, Vincenzo Niola, Sergio Savino and Salvatore Strano .............................. 383

A New Real Time Shape Acquisition with a Laser Scanner: First Test Results
Cesare Rossi, Vincenzo Niola, Sergio Savino and Salvatore Strano .............................. 391

Development of a Modular, Reconfigurable End Effector for the Plastics Industry: A Case
Study
Michael A. Saliba, Andrew Vella Zarb and Jonathan C. Borg ........................................ 399

Perceptions and Problems Related to the Use of Manufacturing Automation in a Small,
Geographically Isolated Economy
Dawn Zammit, Sandro Azzopardi, Michael A. Saliba, and Martin M. Zammit ............... 407

A Research on Fault Detection and Classification of Cyclic Signals Using Spline Regression
and Support Vector Machine
Jonghyuck Park, Sung-Shick Kim, and Jun-Geol Baek ..................................................... 415

The project of a visual programming interface for the Mitsubishi Movemaster RV-M1 robot
Krzysztof Foit and Daniel Reclik ....................................................................................... 423

Automated Deburring System with a 3-Axis Robot for Injection Molded Parts
Seong-Jin Kim, Dong-Won Kim, Gyun-Eui Yang, Eun-Young Heo, Chan-Gie Lee, and F.
Frank Chen ......................................................................................................................... 429

Obstacle and Collision Avoidance in Unstructured and Time-Varying Environments: A Screw
Theory Approach for Manipulators
Carlos R Rocha, Henrique Simas, Daniel Martins, and Altamir Dias ................................ 436

An Experimental Analysis of Bernoulli End Effectors for Contact Manipulation of Irregularly
Shaped Objects
Nebojsa I. Jaksic and Gajendiran Mohandass ................................................................. 444

RFID Systems in Manufacturing SMEs – Benefits and Limitations
Sara Zarei, Simon Hodgson, David Ford and Antony Robson ........................................ 452

A generic Framework for Life-Cycle-Management of heterogenic Building Automation
Systems
Konstantin Krahtov, Sven Rogalski, Daniel Wacker and Jivka Ovtcharova ...................... 461

Improving Process Performance through SMEs-SPC
Mohd Nizam Ab Rahman, Rosmaizura Mohd Zain, Zulkifli Mohd Nopiah, Ahmad Rasdan
Ismail, Jaharah A.Ghani, Baba Md Deros, Norhamidi Muhamad and Suriani Abdul Rahman 469
An Industrial Validation of Artificial Vision Techniques for the Control of Future High Speed Forming Processes
P. Fillatreau, N. Arana, E. Saenz de Argandoña, A. Izagirre, I. Zuriarrain and R. Pop ............ 477

INDUSTRIAL ENGINEERING AND OPERATIONS MANAGEMENT

A Service Target Method for Lead Time Quotation in Assembly Systems
R. Ghasemy Yaghin and S.M.T. Fatemi Ghomi....................................................................... 485

Hybrid metaheuristics for a SDST parallel machines scheduling problem with bi-objectives: makespan and sum of the earliness and tardiness
J. Behnamian, M. Zandieh and S.M.T. Fatemi Ghomi ............................................................. 493

Early/tardy scheduling with sequence-dependent setup times on hybrid flowshop using a novel hybrid metaheuristic
J. Behnamian, S.M.T. Fatemi Ghom and M. Zandieh ............................................................. 501

Setting warm-up period in simulation analysis with autocorrelated data
S.M.T.Fatemi Ghomi and S.Shokuhyar.................................................................................... 509

Efficient Stochastic Hybrid Heuristics for the Multi Depot Vehicle Routing Problem
M. Mirabi, S. M. T. Fatemi Ghomi and F. Jolai ................................................................. 516

Real-time Control of Parts Input in Flexible Production Systems by Analogy Process
Rizauddin Ramli, Hidehiko Yamamoto, and Dzurai dah Abd Wahab ....................................... 524

Quick scan of the production system for disturbances
Paulina Golinska, Marek Fertsch and Pawel Pawlewski ......................................................... 532

Application of Reverse Logistics for Remanufacturing in Electro-Mechanical Industrial Sector
Sahar El Barky, William Ion and Aziz El Sayed ................................................................... 540

A Heuristic Strategy for Stochastic Job-Shop Scheduling with Machine Breakdowns
M. Morady Gohareh, R. Ghasemy Yaghin, and S.M.T. Fatemi Ghomi ................................. 549

An innovative fuzzy time series model for forecasting
S.M.T. Fatemi Ghomi and M.R. Marjani ................................................................................. 557

A new approach in performance evaluation with Artificial Neural Networks
Abdolhamid Safaei Ghadikolaei and Ali Zareshahi ................................................................. 565

Design of a Framework for Inventory Control – Evaluation of Forecasting and Inventory Control Systems
Philip Hedenstierna, Per Hilletofth, and Olli-Pekka Hilmola ................................................. 573

Optimizing Product Placement for Manual Picking in Warehouses
Uday Venkatadri and Sachin Kubasad .................................................................................... 581

An exact approach to minimize the sum of maximum earliness and tardiness in single machine scheduling problem with sequence dependent setup time
Nooshin Nekouiemehr and Ghasem Mosleh ..................................................................... 590
Least Angle Regression Using Orthogonal Basis of Input Variables
Dongil Kim, Jibin Hwang, Sung-Shick Kim, and Jun-Geol Baek ........................................... 598

A Tabu Search approach for pallet transport management problem
Pierpaolo Caricato, Antonio Grieco, and Francesco Nucci ..................................................... 605

Lean Automated Manufacturing – A Panacea or a Potential Path to Ruin?
Hongyi Chen, Richard R. Lindeke and David A. Wyrick ....................................................... 613

A Methodology for Scheduling Part Families in Cellular Manufacturing Systems with
Reconfigurable Machines
Ignacio Eguia, Jesus Racero, Sebastian Lozano, and Fernando Guerrero ............................... 621

Scheduling of production orders with assembly operations and alternatives
Krzysztof Kalinowski ............................................................................................................... 629

Effect of Axial and Radial Velocities on Solids Mass Flow Rate Measurement
Jianyong Zhang and Bin Zhou ................................................................................................ . 636

An Efficient Algorithm Based on TOC to Solve Product Mix Problems
Alireza Rashidi Komijan and Saloomeh Yazdi ................................................................. 645

Case Study of Power System Faults in the Petroleum Development Oman (PDO) Network
Wenping Cao and Nabil Said Al-Kindi ................................................................................... 656

A new strategy for resource-constrained allocation in PERT networks under the condition of
renewable resource
Siamak Baradaran, S. M. T. Fatemi Ghomi, and S.S. Hashemin ........................................... 664

Fuzzy Electre: A Multi-Attribute Method In Fuzzy Environment
Mehdi Iranpoor, S.M.T. Fatemi Ghomi and R. Ghasemy Yaghin ........................................... 672

IT AND INTERNET APPLICATIONS

Driving Method Study and its Implementation for TFT-LCD with High Gray Scale
RAN Feng, XU Meihua and CHEN Zhangjin ........................................................................ 680

A Novel OLED Controller with Fractal Scan Scheme
XU Meihua, RAN Feng, and CHEN Zhangjin ....................................................................... 686

An Improved Technique for Object-based Video Segmentation
Wendan Xu and Donglai Xu ................................................................................................. 694

Critical Infrastructure Protection Security Layer for DNP3 Devices
Todd Mander, Farhad Nabhani and Ivan Lee ....................................................................... 700

Role Based Access Control Infrastructure for Consumer Demand Side Management Internet
Access
Todd Mander, Farhad Nabhani and Ivan Lee ....................................................................... 709
Virtual Enterprise Procurement based on Combinatorial Reverse Auction
Fu-Shiung Hsieh ..................................................................................................................... 717

Integration of Product Design, Process Planning and FMS Control using Neutral Data Representation
Dusan N. Sormaz and Chintan Patel ....................................................................................... 726

LEAN & AGILE MANUFACTURE

Application of the Design of Experiments to Evaluate Production Efficiency in Lean Systems
Steve Davies, Tim Coole, and David Osypiw .......................................................................... 734

Applying Lean Principles to Production Scheduling
Mustafa Ramzi Salman, Roman van der Krogt, James Little and John Geraghty ................. 743

A Fundamental Mindset that Drives Improvements towards Lean Production
Yuji Yamamoto and Monica Bellgran...................................................................................... 751

JIT Performance Measurement in a Specialist Packaging Manufacturing Environment
Y. G. Sandanayake and C. F. Oduoza ................................................................................... 760

Dynamic Workflow Planning in Holonic Manufacturing Systems
Fu-Shiung Hsieh ..................................................................................................................... 769

Leveraging the new SEMI standard to reduce waste and improve flow
Raymond Goss, Carmen Maxim, Diwas Adhikari and Jan Rothe ........................................... 777

Soft Investment: Lean versus Traditional Plants
Horacio Soriano-Meier, Tim Peacock, Paul Forrester, Jose Arturo Garza-Reyes and Luis Rocha-Lona ......................................................................................................................................... 786

MANUFACTURING EDUCATION AND TRAINING

Realistic Manufacturing System Integration Training with an Assembly Plant Prototype Based on Transportation and Warehousing Systems
Julio Garrido Campos ............................................................................................................. 792

Engineering Management Programs For Advanced Manufacturing Technologies
Adji Cissé and David Wyrick ................................................................................................. 800

An intelligent design pedagogy for Engineers.
Keelin Leahy and William Gaughran ...................................................................................... 808

Manufacturing Training Systems in Developing Countries - A Libyan Case Study
A. Afshouk and P. W. Shelton ............................................................................................... 816

Enhancing the Training of Engineering Learners
Sajid Abdullah ......................................................................................................................... 824
MANUFACTURING PROCESSES AND TECHNOLOGY

Effects of Carbon dioxide & Sulphar Contents in Gas Processing Plant
F. Nabhani, Abbas M. Jasim and S. Hodgson ................................................................. 833

Taguchi Method for Optimising the Sintering Parameter of the Metal Injection Moulding (MIM) Compacts

Modelling the nip in a roll press using circular arc principal stress orientation
A.J.Matchett ..................................................................................................................... 843

Use of Mechatronic System to Manufacture Spiral Bevel Gear using 3-Axis CNC Milling Machine
Mohsen Safavi, Mahsa Ghandehary, Reza Abedinzadeh and Mehdi Karimian .................. 850

Aspects of Burr Formation in Bandsaw Teeth Manufactured by Milling Operation
Mohammed Sarwar and Julfikar Haider ........................................................................ 858

Cutting Energy and Wear in Bandsawing Operation When Cutting Ti-6Al-4V Alloy
Fahd N. Khan, Mohammed Sarwar and Martin Persson .................................................. 865

Thermoplastic Starch/LLDPE Blends as a Binder in Metal Injection Molding
Norhamidi Muhamad and Hooman Abolhasani ............................................................. 873

On Commercial off the shelf design of reconfigurable machines tools
K. Mpofu, C. M Kumile, and N. S Tlale ........................................................................... 881

The introduction of a KPI to monitor Spent Foundry Sand disposal and examining and implementing a sustainable alternative.
Martin McKie, Farhad Nabhani, Kirk Bridgewood and Richard McKie ......................... 887

Modelling and Implementing Circular Saw Blade Stone Cutting Processes in STEP-NC
Julio Garrido Campos and Ricardo Marín Martín ............................................................ 896

Ink-Jet Printing System for Manufacturing Conductivity Patterns
Hee-Kwan Lee, Dong-Won Kim, Gyun-Eui Yang, Eun-Young Heo and Jong-Yeong Lee .... 904

Determine the Optimum Settings for Maximum Force (Fmax) on Quad-Flat-Pack’s (QFP’s) using a Wetting Balance Machine.
Robert Woods and Christy Gillick .................................................................................. 912

Understanding the Dynamic Response of a Machine Tool Structure by Finite Element Analysis
Q. Xu and S. Barrans ....................................................................................................... 919

Methodology of soldering process defects management
Frantisek Steiner .......................................................................................................... 925

A review of finite element analysis of metal cutting
Q. Xu and Z. Lu ............................................................................................................. 932
An Intelligent System for Analyzing Welding Defects using Image Retrieval Techniques
Raoul Pascal Pein, Joan Lu, John Birger Stav, Q. Xu, Miro Uran and Luboš Mráz ..............939

A methodology for the determination of foamed polymer contraction rates for the cryogenic manufacture of shoe insoles
V.G. Dhokia, P. Crabtree, S.T.Newman and M.P.Ansell ..........................................................947

A Power Consumption Model for Slot Generation with a CNC Milling Machine
R. Imani Asrai, S. T. Newman and A. Nassehi ........................................................................955

Role based Self-adaptation of a robot DiMS based on system intelligence approach
Kai Salminen, Hasse Nylund and Paul Andersson .................................................................964

MATERIALS TECHNOLOGY

Effects of Chemical Functionalized MWCNTs/Epoxy Matrix Composites on Electrical Conductivity Behavior
Abu Bakar Sulong, NorhamidiMuhamad, Che Husna Azahari, Ahmad Rasdan Ismail and Joohyuk Park .............................................................971

MATERIALS TECHNOLOGY MICRO AND NANO MANUFACTURING

The Feasibility of Printable Power
Huw J Lewis ...................................................................................................................................979

OTHER ALLIED TOPICS

EcoFLEX - Inter-branch Methodology for Flexibility Measurement of Production Systems
Sven Rogalski, Jan Siebel and Jivka Ovtcharovoa ..................................................................987

Beneficial Re-use of Waste Foundry Sand: argument for facilitated Industrial Symbiosis
Kirk Bridgewood and Martin. G. McKie ..............................................................995

Thermal Effects and Pressure Distribution in Regenerative Braking
S. Sarip, A. J. Day, P. Olley and H. S. Qi..............................................................................1003

Stresses in bulk solids in cone hoppers: numerical solutions to the 3- dimensional stress distribution problem, using circular arc geometry
J.C. O’Neill, A.J. Matchett and A.P. Shaw .............................................................................1011

A Bayesian approach for predictive maintenance policy with imperfect monitoring
Giuseppe Curcuru and Giacomo Galante ..............................................................................1019

Robust Maintenance Scheduling for Reliability Maximization
A. Certa, G. Galante, T. Lupo and G. Passannanti ...............................................................1026

A model for a periodic preventive maintenance policy for a series parallel system
A. Certa, G. Galante, T. Lupo, A. Passannanti and G. Passannanti ....................................1034
A Multi-Objective Fleet Routing Optimization for Passenger Shipping among Islands
L. Cannizzaro, M. Enea, C. M. La Fata and G. Passannanti ................................................... 1042

AIDC Technologies and their impact on Management Information Systems
Athanasios Klonis and Farhad Nabhani ................................................................. 1050

Comparison of two tools for the measurement of interfragmentary movement in femoral neck fractures stabilised by cannulated screws
Edward J. Bradley and Farhad Nabhani ................................................................. 1058

Evaluating Residual Life using Failure History of a Machine Tool
B. V. Sudheer Kumar, C. S. P. Ra and A. Neelakanteswara Rao ................................. 1066

Maintenance Management Strategy Development: An Explorative Study in a Large Oil Company
Jalal Ashayeri and Daan F.J. Schraven ........................................................................... 1073

Designing a Network Protocol for Control Systems based on the Hypertext Markup Language (HTML)
Anwar Bashir, Farhad Nabhani and Terry Shaw ........................................................ 1084

RFID-Enabled Real-Time Mass-Customized Production Planning and Scheduling
George Q. Huang, Mike J. Fang, Haili Lu, Q.Y. Dai, WN Liu and Stephen T. Newman .... 1091

Inventory Lot-sizing and Supplier Selection with Shortage Cost in an Assembly System
J. Razmi and G. Assadipour ......................................................................................... 1104

Developing a Mathematical Model to Evaluate and Select the Supplier/s in Presence of Multiple Price Discounts and Budget Limitation
J. Razmi and E. Maghool ............................................................................................ 1109

PRODUCT DESIGN/DESIGN FOR MANUFACTURE/ASSEMBLY

An Investigation of Manufacturing Challenges in Automotive Software Development from a Lean Perspective
Joakim Pernstål and Peter Öhman .............................................................................. 1116

Reusability Enhancement of Automotive Components through Disassembly Study
D.A.Wahab, Lily Amelia, A. R. Ismail, R. Ramli, N. Muhamad and C.H. Che Haron ....... 1124

A CAD System for Extraction of Mating Features in an Assembly
Arun Tom Mathew and C.S.P. Rao ............................................................................. 1132

System Based on Frame to Formalize and (Re)use Functional Knowledge in Functional Modeling of Parts
Francisco das Chagas Mendes dos Santos and Altamir Dias ....................................... 1139

Recognition of Interacting Volumetric Features Using 2D Hints
Dušan N. Šormaz and Chandu Tennety ..................................................................... 1146
Design of Seat Belts: Safety and Comfort for Expectant Mothers
Lily Amelia, D.A. Wahab, S. Abdullah and K. Khalid .......................................................... 1155

PRODUCT DEVELOPMENT, RAPID PROTOTYPING, CONCURRENT ENGINEERING
Ontology-based Data Process for Integrated Data
Xiaomeng Chang, Yanfeng Li, Brandon Brisbon and Janis Terpenny ................................. 1163

New Product Development in a Manufacturing Company – A Challenge for Supply Chain Management
Per Hilletofth, Dag Ericsson, Olli-Pekka Hilmola and Philip Hedenstierna ............... 1169

QUALITY CONTROL
Quality Controls and Testing processes in Machine Tool Realisation
Francesco Aggogeri, Marco Mazzola and Nicola Pellegrini ............................................. 1178

Using Statistical Process Control to Monitor Inventory Accuracy
John R. English, Kyle Huschka, and John M. Miller ..................................................... 1186

QUALITY MANAGEMENT AND TOTAL QUALITY MANAGEMENT
Structuring stakeholders’ requirements concerning intra-logistic facilities in the context of QFD
H.-A. Crostack, S. Klute and R. Refflinghaus .............................................................. 1192

Implementation of a Methodology for the Management of Corrective Actions through Quality Improvement Teams at a Manufacturing Company
Joao C.E. Ferreira and Kleber D. Ristof ........................................................................... 1200

Improving transport productivity within a food distribution SME through a simplified six sigma methodology
Aliireza Shokri, Farhad Nabhani and Nazir Hossien ......................................................... 1208

Requirements for Roll Conveyors and their (Inter-)Dependencies Based on Domain Knowledge
Horst-Artur Crostack, Jonas Mathis and Robert Refflinghaus ........................................ 1218

Characteristics of Successful Kaizen Events: An Empirical Study
Jennifer A. Farris, Eileen M. Van Aken, Toni Doolen, June Worley and Wen-Hsing Liu .... 1226

Developing a Business Improvement Strategy based on self-assessments using the EFQM Model
Luis Rocha-Lona, José Arturo Garza-Reyes, Horacio Soriano-Meier, Tim Peacock and Paul Forrester .................................................................................................................. 1234

Application of Quality Function Deployment to Study Critical Service Quality Characteristics and Performance Measures
Baba Md Deros, Norashikin Rahman, Mohd Nizam Ab. Rahman and Ahmed Husam Said .. 1243
Manufacturing Process Improvement: Perceived Barriers from Adoption of Good Practice by SMEs.
Smith P.G., Oduoza C.F., and Barber K. .......................................................... 1251

SIMULATION, MANUFACTURING SYSTEMS

Modelling Thermal Radiative Properties of Nano scale Multilayer with Incoherent Formulation
Seid Amir Abbas Oloomi ................................................................. 1257

Simulation of service-oriented and distributed manufacturing systems
Nylund Hasse, Salminen Kai, and Andersson Paul H .................................................. 1269

A.O.M. Adeoye, and T. Szécsi ........................................................ 1277

Lifetime Models for Condition Monitoring and Simulation Based RAMS Optimizations
Michael Walther, Alexander Verl, Steffen Nebel and Bernd Bertsche .......................... 1286

Prediction of Operational Performance for Food Manufacturing Systems Using Discrete Event Simulation and Response Surface Methodology
Ayad K. Mohamed, John Geraghty and Saleem Hashmi ........................................ 1294

Analysis of capability of Lean Manufacturing implementation - case study
Anna Sobis, Joanna Oleskow-Szlapka and Marek Fertsch ........................................... 1302

Discrete event simulation for manufacturing system analysis: an industrial case study
A. Gatsou, X. V. Gogouvitis, G.C. Vosniakos, C. Wagenknecht and J.C. Aurich ........... 1309

On the Suitability of the Information Fusion JDL-U Model as a Reference Model for Virtual Product and Production Development
Leo J De Vin ................................................................. 1317

Work center optimization methods for semiconductor manufacturing
Andreas Klemmt, Jan Lange, and Gerald Weigert ........................................... 1325

Strategic Resources Management in Health Care System: The ED Case
Fabio Fruggiero, Alfredo Lambias and Luca Zampognaro .................................... 1333

Solving Complex Crew Allocation Problems in Labour-Intensive Industries Using Genetic Algorithms
Nashwan Dawood, and Ammar Al-Bazi ......................................................... 1340

Optimisation of Operation Sequences in Flexible Manufacturing Cells using Virtual Manufacturing Tools
Magnus Holm, Josefine Doverborn, Amos Ng and Leo De Vin ........................... 1348

A STEP Compliant System for Manufacturing of Composites in the Aerospace Industry
Monitoring and Control Instrumentation of Gravity Crude Separator
Mohamed Sawessi, Farhad Nabhani and Ian French .............................................................. 1365

SUPPLY CHAIN MANAGEMENT, BUSINESS STRATEGY

Supplier Development Practice: Arising the problems of upstream delivery for a Food Distribution SME in the UK
Alireza Shokri, Farhad Nabhani and Simon Hodgson .............................................................. 1375

Deployment of ERP for Supply Chain Excellence in a Manufacturing Organisation
Stephen Williams, Zeynel Badak, and Chris Thompson ....................................................... 1386

Designing supply contract to coordinate supplier-manufacturer interactions, when manufacturer's leftover products have stochastic demand
Iman Nossohi and Ali Shahandeh Nookabadi ....................................................................... 1395

Analysing turbulences in sequence fixed production systems according to the pearl chain production planning and control system (ppcs)
Danny Szendrei, Katja Unger, Tobias Teich & Tim Neumann ............................................. 1403

Alternative strategies for Vendor Managed Inventory systems in single-vendor/multi-buyer supply environments
M. Falivene, R. Iannone, S. Miranda and S. Riemma .............................................................. 1411

Development of an Online Supplier Selection System for a Turbine Manufacturing Company
L.Siva Rama Krishna, G.Ranga Janardhan and C.S.P Rao ..................................................... 1419

Handling of Time, Complexity and Dynamics within Supply Chains
Thomas Burghardt, Hendrik Jähn, Sascha Häckel and Sascha Lemke ................................... 1427

Supply Chain Management Constraints: Telecom Industry Case Study
Fatiha Naoui ....................................................................................................................... 1435

How Quality Management could improve the Supply Chain performance of SMES
Paul Eric Dossou and Philip Mitchell .................................................................................... 1442

Establishing a framework for building a SC performance system
Alberto de la Calle and Esther Alvarez ................................................................................. 1450

Optimization of Multi-Echelon Supply Chain Networks using Quantum Inspired Particle Swarm methodology
M Bachlaus, F F Chen, W.H.A. Wang and D.W Kim ........................................................... 1458

Exceptions management in the supply chain
Esther Álvarez, Fernando Diaz and Miguel A. Larrinaga ...................................................... 1467

Comprehensive performance analysis of cross-docking system considering transportation consolidation
Yifan Wu, Ming Dong and F. Frank Chen ............................................................................. 1476
A System Dynamics Based Evaluation of Collaboration Strategies to Manage Demand Gap in New High-tech Product Diffusion
Xumei Yuan, Ling Shen, Jalal Ashayeri ................................................................. 1481

**SUPPLY CHAIN STRATEGIES**

Joint Replenishment and Supplier Selection using Particle Swarm Optimization
Y.-C. Wang and W.-J. Li .......................................................................................... 1491

Going Green in Supply Chain Management as a Prospective Improvement
S. Rania A.M. Shamah ............................................................................................. 1498

Tool support for industrial engineering in small enterprises
Egon Müller, Sebastian Horbach and Jörg Ackermann ........................................... 1512

Radio Frequency Identification (RFID) in Manufacturing Enterprises: From Data to Information
Can Saygin and Jagannathan Sarangapani ............................................................ 1519

Dynamic Analysis of a Simplified Centralised Supply Chain
Ian French, Petia Sice and Mu Niu .......................................................................... 1527
Engineering for Sustainability

Prof Graham Hillier, BMet, PhD, MBA, FIMMM, FRSA

Director of Low Carbon Energy
Centre for Process Innovation
B210, Wilton Centre
Redcar
TS10 4RF

ABSTRACT:

There have always been three main principles behind successful business. They are:

- Economic factors – that create the wealth to do things and continue to do them
- Environmental, or natural resource factors – the impact we have on the resources available to us on the Earth
- And Societal factors – that people have healthy, happy and full lives.

The world is facing ever-increasing pressures on its natural resources. Population in 2003 was 6 billion and the UN forecasts that there will be over 8 billion people by 2030. All of these people need homes and services and many aspire to a developed world standard of living. In addition to the rapid growth in population and basic need, energy prices are rising and security of supply is becoming a major issue. There is also an increasing belief that humans are causing severe damage to the global environment.

Creating a balance between the economic, environmental and societal factors is what is now termed sustainable development. The modern engineer has a key role in helping the global community respond by developing and introducing technologies that deliver a more balanced approach.

This paper looks at how the engineer’s role will change to respond to these needs by taking a new look at problems and processes. It discusses how engineering technology can help make better use of the planet’s natural resources by using novel engineering systems and processes. It also considers the interaction between engineering, the built environment and social impacts.

Corresponding Author: E Mail: graham.hillier@uk-cpi.com
Successful Continuous Improvement – How to make and sustain improvements

Keith Copeland MBE

Senior Engineer NMW and V-UP Office
Nissan Motor Manufacturing (UK) Limited
Sunderland Plant

ABSTRACT
As the global economy continues to free fall at a staggering rate manufacturers are facing ever increasing pressures not just to remain competitive but to survive. Realising the urgent need for increased productivity whilst reducing costs, many have been forced to adopt numerous Improvement Methodologies. Research has found that 72% of these methodologies are failing to deliver expected improvements with companies failing to align improvement strategies with their actual business needs; whilst other issues included a failure to sustain improvements beyond a 12 month period.

This session will discuss the processes used by Nissan in their approach to continuous improvement and how they have coped with the current market difficulties they face. The session will also discuss research from the Nissan supplier base and wider manufacturing sectors looking at the reasons for failure and poor sustainability looking at how companies identify issues within their organisations and how they select the appropriate tools or methodologies to address these issues.
RFID-Enabled Wireless Manufacturing Execution Platform

Professor George Q. Huang

Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong,
Pokfulam Road, Hong Kong, P. R. China

ABSTRACT

This talk discusses how RFID (Radio Frequency Identification) technology is used to develop real-time and easy to use information platform for reconfigurable manufacturing enterprises. RFID devices are deployed at the shop-floor for collecting real-time manufacturing data from manufacturing facilities, cells and lines. The platform interacts and interoperates seamlessly with enterprise information systems such as ERP (Enterprise Resource Planning) and WMS (Warehouse Management Systems). The adoption of the platform improves productivity and product quality while reducing costs and reworks through real-time information traceability and visibility. The enterprise competitiveness is ultimately improved through better market and engineering responsiveness. The talk is divided into three parts. The first part is mainly concerned with how RFID devices are deployed to value-adding manufacturing objects and facilities such as machine tools, pallets, materials, tools, operators and Work In Progress (WIP) items. The second part of the talk will deal with how smart objects are combined with production processes and operations in order to achieve the seamless integration between shop-floor and enterprise information systems. Finally, visibility and traceability facilities are presented for different types of users such as shop-floor supervisors, process planners, production planner and schedulers, equipment maintenance engineers, and line operators. These facilities will allow them to visualize and monitor the status and provide real-time decision supports.

Corresponding Author: gqhuang@hku.hk
Bringing economic success through productivity and engagement

Dr Stan Higgins
CEO
North East Process Industry Cluster
Room H224, Wilton Centre
Redcar, TS10 4RF

ABSTRACT:

Dr Higgins will put the Process Industry is put into its economic context in both the United Kingdom and the North East Region. He will explain how economic clusters work strategically today and can become the economic powerhouse for a region. How areas for public and private sector collaboration can be identified to bring about transformational performance of a sector and for participating companies. Key is a focus on the whole supply chain, productivity and resource efficiency as components of competitive advantage for a region, the cluster as a whole and for individual businesses. Dr Higgins will identify key success factors that must be acted upon by the industry and its stakeholders to bring success.

Corresponding Author: Email: stan.higgins@nepic.co.uk
Where lies the Factory of the Future?

David W Russell, PhD FBCS, FIET, FI MechE

Penn State Great Valley
Malvern, PA 19355, USA

ABSTRACT:
The Factory of the Future will regretfully evolve slowly as the number of new manufacturing facilities declines in the Western hemisphere and increases in Asia. The call for sustainability often falls on ears defeaned by short term commercial advantage. Undeveloped and developing nations are striving to reach a par with the so-called developed nations – largely in the west – at whose door, along with power stations, much of the environmental problems can be laid.

The emergence of the Knowledge Economy has brought about the reality of worldwide access to information and potential markets; ubiquitous computing will ensure that future generations will be as totally blasé and dependent on IT as previous ones were on printers and FAX machines! What is changing at an incredibly rapid pace is the universal level of dependency on convenience, service, and automation that is based on information technology.

As world societies adopt the consumerism and service-centric lifestyle so prevalent in the west, there will continue to be a need for low cost, high quality and hopefully eco-friendly, technocentric products, but how and where will they be manufactured?

This paper attempts looks at some existing, and future, technologies that seem to be very promising in assisting in the manufacture of better products, more efficiently while solving many of the challenges that corporations must face in implementing them, and looks at the possibilities of the next 20 years and beyond.

Corresponding Author: Tel: +1 610-648-3233; Email: drussell@psu.edu
Framework of Grinding Process Modeling and Simulation Based on Microscopic Interaction Analysis

Xuekun Li*, Yiming Rong

Computer-Aided-Manufacturing Lab, Manufacturing Engineering Program
Worcester Polytechnic Institute
Worcester, MA, 01609, US

ABSTRACT:

This paper describes a framework of grinding process modeling to understand grinding fundamentals and design grinding processes with predictive performance. The model regards grinding process as a time dependent process and an integration of microscopic interactions in the wheel-workpiece contact zone, including cutting, plowing, and sliding as well as other frictional interactions. And the grinding process control and design are in fact to manage and balance all these interactions. The fundamentals of microscopic interactions are analyzed and correlated to the grinding process input parameters and performance output.

1. INTRODUCTION

Grinding is a special machining process with large numbers of parameters influencing each other, which can be considered as a process where the thousands of irregular cutting edges interact with the workpiece at a high speed simultaneously. When perceived in this manner the grinding processes appear to be complex and managed only as a black art. However, irrespective of the choice of variables in the input categories, for every grinding process it is possible to visualize the basic microscopic wheel-workpiece interactions in terms of grain-workpiece interaction, bond-workpiece interaction, chip-workpiece interaction, and chip-bond interaction in Figure 1. Small changes in these 6 modes can have dramatic changes in the output of the system. Hence every abrasive machining process control is an effort to balance between cutting (surface generation) and tribological interactions of plowing/sliding while eliminating all the other frictional interactions [1]. And proper measurement and analysis immensely help application engineers in such strategic management of the grinding processes.

Moreover, grinding processes exhibits a strong time dependent characteristic, which is a combination of all microscopic interaction changing as a function of time. In industry where grinding power signal is widely measured for process monitoring, while the wheels get worn, loading, or glazing, the power curve will show a steady and gradual change as in Figure 2. Superimposing the power profiles of cycle 5 and cycle 1 makes the change visible, as shown in Figure 3. Within one individual grinding cycle which consists of several segments: rough, semi-finish, finish, spark out, etc. the MRR-Power draw can be obtained by curve fitting into a straight line. And the MRR-Power draw change from cycle 1 to cycle 5 tells the “inside story” of the grinding process. An in-depth analysis of

* Xuekun Li: Tel.: (0001) 508-831-6167; Fax: (0001) 508-831-6412; E-mail: xuekunli@wpi.edu
the MRR-Power draw, in Figure 3, leads to the decomposition of the power in terms of threshold and cutting components and other time dependent components. Each one of these components in turn is associated with specific aspects of the microscopic interactions, as well as the wheel properties alteration leading to such interactions. Figure 3 indicates a qualitative understanding of the MRR-Power draw change, which could predict grinding wheel surface conditions but still insufficient in providing an explicit solution for grinding optimization. And quantification of this power curve superimposition and MRR-Power draw in terms of the microscopic interactions and their change is a key aspect of managing modern grinding processes.

Figure 2: Grinding power change in typical OD grinding processes

(a) Superimposition of grinding power curves   (b) Qualitative decomposition of a time dependent grinding power curve

Figure 3: Superimposition of grinding power signals and its correlation with microscopic interaction modes

2. LITERATURE REVIEW

The literature on grinding process modeling is rather extensive and it would not be possible to cover it in any detail. As grinding force contributes almost all aspects of grinding process output, only pertinent literature dealing with force modeling is covered. S. Malkin [2] decomposed the grinding force into chip formation force, plowing force, and sliding force. The chip formation energy was related with the melting energy for iron, which was about $u_{ch}=13.8 \text{ J/mm}^3$. And the tangential plowing force per unit width was estimated to be 1N/mm for steels. Sliding was associated with rubbing of dulled flattened areas on the abrasive grain tips (wear flats) against the workpiece surface. Thus the grinding force was a summation of all these components. Inasaki [3] measured the cutting edges by counting the peak points on wheel surface. The cross-section area was automatically calculated in his simulation software for calculating the force acting on each single grain. The integration of force on all cutting edges gave the grinding force in global scale. Chen and Rowe [4] modeled a grinding wheel surface with statistical methods. And the single grain cutting force was regarded comparable to indenter-specimen interaction in Brinell hardness test in the absence of friction. The kinematic simulation generated the active grain number and cutting force on each grain. Badger and Torrance [5] used both Challen and Oxley’s 2D plane-strain slip-line field theory and Williams and Xie’s 3D pyramid-shaped asperity model to calculate grinding force on each single grain. Hou [6] incorporated the random nature of grain distribution into their work. The dynamic grinding force was formulated as the convolution of a single-grit force and the grit density function. All the literatures suggest the primary modules for a grinding process model should include a grinding wheel model, kinematics model, and single grain cutting model. However, the chip-workpiece, chip-bond, bond-workpiece interface and the time dependent properties of grinding process are
overlooked in current research. For such reasons, a grinding process model is in demand to quantify the significance and change of each single microscopic interaction as a function of time.

3. **Structure of the Grinding Process Model**

   ![Figure 4: Framework of the Grinding Process Model](image)

   The understanding of grinding processes in terms of time dependent microscopic interactions indeed shed light on the principles for the modeling. First of all, the change of the wheel surface conditions is the fundamental cause of the time dependent behavior of grinding. Therefore, a grinding wheel model is a prerequisite for grinding process modeling, through which the microscopic wheel properties and their change as a function of time can be presented. Secondly, in order to specify and quantify the 6 interaction modes, some criteria should be deduced for the identification and modeling of each mode. The output of each single interaction mode, such as force (or power) consumption, heat source generation, should be correlated with the input parameters in the microscopic interaction model. Thirdly, to determine the input data for microscopic interaction model, a kinematics model is necessary, from which the contacting condition of each interaction can be specified. Meanwhile, in the kinematics model, the integration of the output force (or power) provides the grinding force (or temperature) in macro-scale, and the heat source that is discrete in nature can be deducted from the microscopic force. In total, the time dependent grinding process modeling can be broken down into 3 levels in Figure 4:

1. A grinding wheel model, which describes both surface topographical and mechanical properties.
2. The microscopic interaction model, which serves to categorize and quantify each microscopic mode.
3. A kinematics model, which provides the contact condition for each mode and integrates the output data from the microscopic interaction model.

### 3.1. Grinding Wheel Model

Currently, there exist two popular approaches for grinding wheel modeling. One approach is to use pure mathematical methods to simulate the wheel surface. Alternatively, the other is to apply statistical approaches to solve this problem considering random distribution of grains on the grinding wheel surface [7]. However, the pore volume and the mechanical properties of the wheel, which are critical to the grinding process, cannot be obtained from all the reported methods.
In order to provide a virtual grinding wheel morphology which is equivalent to a real product in terms of topography feature as well as mechanical features, a through-the-process modeling method is proposed. Our idea is to utilize mathematical methods to intimate each wheel fabrication step, from raw material mixing to final wheel dressing. Not only the composition of wheel, such as grain size, grain shape, grain fraction, and bond fraction, but also the mechanics, and bond material diffusion during wheel firing are considered. After dressing simulation, the virtual wheel surface should bear resemblance with the real products in terms of static grain count, protrusion height, effective pore volume, and local wheel hardness, as indicated in Figure 5.

Figure 5: Framework (a) and procedure (b) of grinding wheel modeling

3.2. MICROSCOPIC INTERACTION MODEL

**Grain-Workpiece Interaction: Cutting, Plowing**

Cutting and plowing are the most fundamental interactions in grinding, which modify the workpiece surface directly and dominate the material removal efficiency. Moreover, the chip generation in microscopic cutting also contributes to the chip/bond interface and chip/workpiece interface. Meanwhile, the surface modification by one grain influences the material removal of the successive grain, as indicated in Figure 7 [8]. Therefore, mechanisms of cutting and plowing at microscopic level should be established for comprehensive understanding of grinding. Although a number of grinding experiments with a single abrasive grain were conducted, it is still quite intricate to establish the mechanisms in 3D due to the measurement difficulties of force, temperature and workpiece material deformation. On the contrary, Figure 7 indicates finite element models and packaged FEM software are capable to describe metal cutting processes explicitly. Therefore, there are great possibilities that finite element modeling can
be applied to investigate the single grain material removal under a wide range of grinding conditions. This can clearly quantify the force (or energy) consumption, chip generation mechanism, as well as localized material deformation, which are difficult to acquire based on only the common sense of grinding and single grain test. In this research, a commercialized FEM software package AdvantEdge™, which incorporates the thermo-mechanical properties of material, is employed for single grain material removal understanding.

In grinding, an abrasive grain can be considered to be an inverse cone shape or pyramid in Figure 8(a), and it may vary from different wheel specifications. During grinding, one abrasive grain seldom fully engaged with the workpiece, instead it will only contact partially with the workpiece, as indicated in Figure 7. Then, at the moment \( t \), the force consumption can be expressed as a product of specific cutting force times the grain-workpiece engagement cross-section area \( A(t) \) [3]. In addition, Figure 8 indicates that the specific cutting force is no longer a constant according to micro cutting theory, which should be described as a function of grain-workpiece engagement depth \( d \). The side flow formation for each single grain in cutting and plowing are also need to be considered, as the side flow geometry will affect the material removal of the successive grain. The side flow shape is considered to be a sphere cap, which is characterized by the width \( b \) and height \( h \). The chip volume generated during a time interval \( \Delta t \) can be derived for loading force calculation, which will be explained later in this paper. Through the FEM simulation, the force consumption, side flow geometry, and chip volume generation for each grain-workpiece contact couple can be deducted from following equations, as indicated in Figure 8.

\[
\begin{align*}
F &= \text{Specific Force} \cdot A(t) \\
\text{Specific Force} &= f(d) \\
F &= f(d) \cdot A(t) \\
b &= f(v, d, \text{grain shape}, t) \\
h &= f(v, d, \text{grain shape}, t) \\
\text{chip volume} &= f(b, h, \text{grain shape}, t) \cdot v \cdot \Delta t
\end{align*}
\]

**Grain-Workpiece Interaction: Sliding**

Figure 9 shows the grain-workpiece sliding occurrence when wear flat is developed on the grain tips. The area of wear flat gradually increases with grinding time and the rate of growth depends on the grain-workpiece combination, grinding parameters and the environment parameters. With the growth of wear flat, both the tangential and the
normal forces increase, thus resulting in further increase in grinding energy consumption, grinding temperature and thermal damage [2]. The total wear flat area, which is the summation of wear flat area on each grain tip, is found to be in linear relationship with the grinding force [9]. Considering the proportional relationship between the grain-workpiece friction force and grain-workpiece contact area, Equation 7 shows the sliding force (power) consumption.

\[ F_{G-W} = \mu_{G-W} \cdot \tilde{p} \cdot A_{G-W} \]  

Where, \( \mu_{G-W} \) is the friction coefficient between grain and workpiece material;
\( \tilde{p} \) is the pressure on the interface, which can be derived from previous cutting simulation;
\( A_{G-W} \) is the area of grain and workpiece contact.

According to the tool wear model developed by Usui [10], the wear flat development during grinding can be estimated as a function of time as in Equation 8.

\[ \frac{dA_{G-W}}{dt} = C \cdot \delta_n \cdot V_s \cdot \exp\left(-\frac{\lambda}{\theta}\right) \]  

Therefore, Equation 9 shows the calculation of the sliding force, which is equivalent to the glazing force.

\[ F_{G-W} = \mu_{G-W} \cdot \tilde{p} \cdot C \cdot \delta_n \cdot V_s \cdot \exp\left(-\frac{\lambda}{\theta}\right) \cdot dt \]  

![Figure 9: Grain-Workpiece Sliding in Grinding](image)

**Chip-bond and Chip-workpiece Interaction**

Chip-bond and chip-workpiece friction, absent in any other cutting processes, is one key area that differs grinding from cutting. These interactions usually happen associated with wheel loading, which is one of the undesirable phenomena in grinding. As wheel loading increase, additional energy is consumed due to the increased friction. Loading is the criterion for wheel dressing with fine grinding in 15% or even higher [11]. Wheel loading can be defined as accumulation or adhesion of grinding chips at the inter-grain space, as indicated in Figure 10.

Therefore, it can be inferred that the relationship between chip volume generated by the grain and pore volume in front of the grain is likely to influence the tendency for loading. Aside from the chemical effect, this can clearly be avoided by appropriate choice of grinding and wheel parameters. One possible solution without reducing grinding efficiency is to use wheel with high porosity or improved porosity shape for easier chip flow [12]. For this understanding, the loading force (power) consumption can be modeled as a function of ratio between leftover chip volume and pore volume, which can be described as:

\[ F_{B-W} = \mu_{avg} \cdot \tilde{p}_{chip} \cdot \left(\frac{V_{chip}}{V_{pore}}\right) \]  

\[ V_{chip} = \emptyset \cdot \int_{t_1}^{t_2} chip \, volume \, dt = \emptyset \cdot \int_{t_1}^{t_2} f(b, h, grain \, shape, t) \cdot dt \]  

where, \( \mu_{avg} \) is the average friction coefficient for chip-bond and chip-workpiece interface;
\( \tilde{p}_{chip} \) is the pressure in the interface, which is associated with the wheel hardness;
\( V_{pore} \) is the effective pore volume in front of a contacting grain;
$V_{\text{chip}}$ is the chip volume remaining in the pore in front of the grain, which is the summation of chip leftover from previous cuts and the chip generated by current cut.

$\phi$ is related with the cleaning efficiency of the coolant. Smaller $\phi$ is for higher pressure coolant, which can help evacuate the chip clogging in the pore more effectively while reduce $V_{\text{chip}}$ and the loading force consequently.

**Bond-workpiece Interaction**

Bond-workpiece is another key area that differs grinding from cutting. As grain wears or breaks down from the bond material, some of bond material rubs against the workpiece material consuming energy and as a result raising the specific energy requirement and heat generation, as shown in Figure 11. The bond-workpiece friction seems insignificant in conventional grinding, however, would play an important role in superabrasive machining where the metal bond is used as bonding agent [13]. The calculation of friction force between bond and workpiece material is similar to grain-workpiece sliding, the proportional relationship between the bond-workpiece friction force and bond-workpiece contact area still applies. Therefore, the bond-workpiece friction force can be expresses as:

$$F_{B-W} = \mu_{B-W} \cdot \bar{p} \cdot A_{B-W}(t)$$  \hspace{1cm} (12)

Where, $\mu_{B-W}$ is the friction coefficient between bond and workpiece material;

$\bar{p}$ is the strength of the bond, which is associated with the wheel hardness;

$A_{B-W}$ is the contact area of bond and workpiece contact.

3.3. Kinematics Model

The kinematics model serves to simulate the wheel moving against the workpiece under the specified grinding condition, which provides the number of contacting grains, contact cross-section area for each grain, and contact area of bond-workpiece interface, as shown in Figure 11. To determine the time dependent properties in grinding, the kinematics simulation is performed as an iteration procedure. During each iteration interval $\Delta t$, the wheel surface move relatively respect to the workpiece complying with the grinding parameters. The material removed by the simulated wheel sample can be regarded equivalent to a triangular shape in cross-section, as indicated in Figure 11. The height $h_m$ and length $l_c$ are corresponding to the maximum chip thickness and contact arc length, respectively, which can be derived analytically from existing literatures [2]. Both wheel and workpiece surface conditions are imported from previous simulation step, based on which the wheel-workpiece contact mode is calculated. Through calling the microscopic interaction models, the force consumption for each single contact couple can be obtained. Integration of the all the microscopic force gives the grinding force (power) consumption in global scale, which also gives a discrete heat source moving along the workpiece surface for future grinding temperature calculation.
Grinding Force = Cutting force + Plowing force + Sliding force + Loading force + Bond_Work force \hspace{1cm} (13)

4. Potential Applications

As a whole, the grinding process model enables the decomposition of grinding force components as a function of time. Since grinding force components are associated with the grinding wheel surface property change, the grinding process model helps predict the wheel surface evolution for grinding diagnoses and optimization. Furthermore, the grinding process model enables the calculation of grinding force (or power) for a cycle, which consists of several segments: rough, semi-finish, finish, spark out, etc. Therefore, the grinding process design can be carried out proactively while eliminating ‘trial and error’. In addition, the grinding wheel model itself can be used to guide the development and optimization of grinding wheels.

5. Conclusions and Future Research

The grinding process model framework based on the understanding and analysis of the time dependent microscopic interactions is developed, and the modules that constitute this model are described in detail. This model enables a quantitative description of the “inside story” in the grinding zone through separation and quantification of the 6 microscopic interaction modes. Further developments of this model include incorporation of different grinding processes, such as slot grinding, and profile grinding, application of more efficient computation algorithms.

Acknowledgements

This research is partially supported by Worcester Polytechnic Institute Fellowship and the open project from State Key Laboratory of Digital Manufacturing Equipment and Technology (Huazhong University of Science and Technology, China). The authors would like to thank Surface Preparation Technologies Group at Saint-Gobain High-Performance Materials for the helpful discussions and suggestions.

References

Clean Technology approach for use of Micro-bioreactor in Bioprocess Engineering

Pattanathu K.S.M. Rahman*, Godfrey Pasirayi, Vincent Auger, and Zulf Ali

School of Science and Technology,
University of Teesside, Middlesbrough –TS1 3BA,
Cleveland, United Kingdom.

ABSTRACT

Low-cost polymer polytetrafluoroethylene (PTFE) was used to fabricate a 1.5ml micro-bioreactor for the production of microbial surfactant (detergent). The choice of PTFE was attributed to the high permeability to gases, low cost and the ease with which it can be sterilized under steam or chemicals. A model microbe Pseudomonas aeruginosa DS10-129 was used in this study. The progress of fermentation/bioprocess was monitored by comparing the growth of the organism in a micro-bioreactor, conventional bioreactor and shake flask methods. Under the micro-bioreactor conditions used Pseudomonas aeruginosa DS10-129 produced 106g/ml of effective biosurfactants within 16h of incubation that reduced the surface tension of the culture medium to 27.9mN/m and emulsified kerosene by 71.30% that showed promising applications in enhanced oil recovery. A green pigment, pyocyanin was produced during the exponential phase, while biosurfactant production was observed during stationary phase of growth. The pigment showed antibiotic effect on various bacteria like Bacillus subtilis, Staphylococcus aureus, Staphylococcus epidermidis, Pseudomonas teessidea and Pseudomonas clemancea. The results obtained shows that the growth pattern and the production of biosurfactant and pyocyanin in the micro-bioreactor are comparable to those obtained by using the shake flask and the conventional fermentation (batch fermenter) methods. Hence the low-cost micro-bioreactor can perform the same functionality as the conventional bioreactors.

1. INTRODUCTION

The development of micro-total analytical system (µTAS) has seen researchers focussing on scaling down expensive grandiose equipment to miniature scale with the overall aim of improving throughput, enabling fast analysis, optimising operations and cutting down on resources. Currently, microfluidic systems are being employed to study diverse processes such as bioprocessing developments, chemical synthesis, clinical diagnostics and genetic engineering.

Research has shown that microfluidic systems are a promising technology that has the potential to improve analytical performance, reduced laboratory safety requirements and costs, shorter analyses time and above all reduced reagent use [1]. Bioprocessing operations such as fermentation has been widely used to realize a number of products of economic importance such as antibiotics, vaccines, therapeutic proteins and many others. However the current mode of operation for these processes is the use of conventional stirred tank bioreactors with typical volumes between 0.5 and 10 l and shakes flasks [2].

These conventional systems have some limitations which includes among other the following: (i) conventional bioreactors are bulky; they consume a lot of resources thereby making bioprocessing an expensive undertaking. (ii) process steps are labour intensive to operate and maintain i.e., sterilizing, cleaning, assembling and dissembling of the bioreactor components, (iii) generation of large volumes of metabolic waste, which may be toxic to natural ecosystems and costly to manage, (iv) sampling process is prone to contamination due to the magnitude of the manipulations that are made, and may result in changing the experimental conditions and (vi) systems are not amenable to performance of experiments in parallel. To mitigate these constraints novel studies on designing and fabrication of miniaturized bioreactors have been documented.
To date a number of different microbioreactors have been reported in literature [3], [4], [5], [6]. These reports have shown that low volume, disposable, microbioreactors with parallel processing operations, and integrated real time measurements are a promising tool in high throughput bioprocessing developments. For example previous work by [3],[6] have shown that microbioreactors provide unique environments, which are ideal for rapid and efficient biosynthesis due to the scale dependence of thermal and mass transfers.

A portable anaerobic microbioreactor to study the optimum growth conditions for the methanogen, *Methanoseta concilii*, a methane-producing obligate archaeabacterium was developed [7]. This work demonstrated the ability of *M. concilii* to grow in the bioreactor microfluidic channels under optimum conditions of temperature, nutrients and pH. Recently microbioreactor arrays developed for controlling cellular environments in human embryonic stem cells [5]. The system uses a simple and novel practical system that couples a microfluidic platform with an array of microbioreactors. A 150μL, well mixed polymer based microbioreactor with integrated optical measurements also developed [8]. In this bioreactor mixing was achieved by magnetic stir bar and fluorescent sensors were integrated for online measurement of pH and dissolved oxygen. It was shown that the behaviour of *E.coli* fermentation in a microbioreactor was identical to the results obtained in a conventional bench scale bioreactor. In their sterling work [3] pioneered and demonstrated online measurements of optical density (OD), pH, and dissolved oxygen in a 1L fermentor. In another development a miniature bioreactor with a working volume of 6mL has also been reported [9]. Oxygenation was achieved by sparging while mixing was enhanced by an impeller.

*Diluzio et al* used microfluidics to study the interactions of motile cells of bacteria with surfaces[10]. They observed that hydrodynamic interactions between cells and the walls of the channels cause the cells to swim preferably on one side of the channel. On the other hand, [11] and the Quake group developed a multi-compartment microfluidic bioreactor composed of PDMS peristaltic pumps and microvalves. The device was fabricated using soft lithographic techniques and was used for sustaining the long term controlled growth of a small population of *E. coli* in the microfluidic loop based bioreactor.

A detailed review [12] highlights the current practices and future opportunities in miniature bioreactors. The review focuses on the use of microbioreactors and their application as an effective tool in bioprocessing, media developments and strain improvement to process optimisation. In view of these developments it can be seen that microbioreactors have the potential to replace labour intensive shake flask and conventional bench bioreactor methods in the screening phase of biological experiments.

In this work we describe the design of a batch fed microbioreactor made from Polytetrafluoroethylene (PTFE) tubes of 0.8mm in diameter with a working volume of 1.5ml. The choice of using PTFE was attributed to its permeability to gases, less costs and the ease with which it can be sterilized. To assess the performance and effectiveness of the microbioreactor, we used *P. aeruginosa* DS10-129 sampled from sites contaminated with diesel as the model microorganism [13] [14] [15]. In this work, we report the fermentation of *P. aeruginosa* in the microbioreactor, fermenter and shake flask methods in terms of growth and productions of extracellular metabolites (pyocyanin and biosurfactants).

2. MATERIALS AND METHODS

2.1 MICROBIOREACTOR DESIGN

The microbioreactor was made from a synthetic polymer polytetrafluoroethylene (PTFE). A PTFE tube (length 3 m 0.8 mm diameter) was made into coils of diameter 4.2 cm to give a microbioreactor of working volume 1.5 ml (Fig 1).

2.2 MICROORGANISM AND CULTURE MEDIUM CONDITIONS

*Pseudomonas aeruginosa* DS10-129 (AM419153) sampled from diesel contaminated sites [13] [14] [15] was used for all the experiments. To promote pyocyanin production, DS10-129 was cultured in glycerol supplemented nutrient broth (GSNB) medium. The organism was regularly resuscitated in Glycerol Supplemented Nutrient Broth (GSNB) medium and working cultures were sub cultured on nutrient agar plates and incubated at 30°C for 18-24h. The composition of the Glycerol Supplemented Nutrient Broth (GSNB) medium was (g/l): Lab Lemco Powder 1.0, yeast extract 2.0, Peptone 5.0, Sodium Chloride 5.0, 30ml glycerol and 1litre of deionized water. The pH of the medium was adjusted to 6.5±0.2. Single colonies of DS10-129 taken from 18-24 hour cultures were used to prepare
inocula in sterile GSNB medium. The inoculated media was incubated in a rotary shaker set at 30°C and 150 rpm. When the optical density of the culture medium was between 0.8 -1.0, it was used to inoculate the GSNB media to give a starting optical density of 0.1 at 600nm. The inoculated media was also used to for the microbioreactor experiments.

2.3 **Fermentation in the Microbioreactor**

By using aseptic techniques, the sterile GSNB medium was inoculated with DS10-129 suspension to give a starting OD$_{600}$ of 0.123, equivalent to approx. 5.7 x 10$^8$ cells/ml. The inoculated media was withdrawn with a 5 ml syringe and mounted on a syringe pump (Razel Scientific, USA) and was used to inoculate the microbioreactor as shown in Fig. 1. After the inoculation procedure, the inlet and outlet ports of the microbioreactor were each connected to a 0.2 µm Whatman sterile syringe filter. This was to prevent contamination and allow diffusion of gases in and out of the microbioreactor. A total of three microbioreactors were prepared and inoculated in the same manner and allowed to run in parallel. The microbioreactors were incubated at 30°C with shaking at 150 rpm for 24h.

At predetermined intervals all the three microbioreactors had their contents sacrificed and monitored for cell growth by measuring the OD$_{600}$ of the metabolised GSNB medium. The metabolites produced by the organism were collected by centrifuging the culture broth at 8000g at 4°C for 20 min to remove bacteria. The supernatant was filtered through a 0.2 µm syringe filter, protected from light and kept refrigerated for subsequent analyses.

2.4 **Fermentation in the Fermenter and Shake-flask**

Parallel batch fermentations of DS10-129 to monitor the production of pyocyanins and rhamnolipids were carried out in a fermenter and shake-flask under agitation for 120 h. A 7.5 l bioreactor (BioFlo 110, New Brunswick Scientific, USA) was filled with 3 l sterile GSNB medium and inoculated with a suspension of DS10-129 with optical density 0.8–1 to give a starting OD$_{600}$ of 0.123. The temperature, pH and agitation speed were maintained at 30°C, 7 and 200 rpm, respectively, but the levels of dissolved O$_2$ were not monitored. Samples of the metabolised culture medium were periodically withdrawn for OD$_{600}$ measurements and subsequent down-stream processing as described for the microbioreactor.

Similarly, a 250 ml conical flask containing 100 ml sterile GSNB medium was inoculated with DS10-129 and incubated at 30°C with shaking at 150 rpm. Samples were withdrawn periodically and treated as described for the other systems.

---

**FIG 1.** Schematic set up of the microbioreactor. The GSNB and the inoculum are introduced into the microbioreactor by a syringe pump as a batch feed. Sterile 2µm Whatman syringe filter are fitted at the inlet and outlet (when the syringe pump has been disconnected) of the microbioreactor. The purpose of the filters is to allow diffusion of gases in and out of the microbioreactor.
2.5 Pyocyanin Production

The presence of pyocyanins was detected by visual observation of a green pigmentation produced by the organism in the nutrient broth culture.

2.6 Inhibitory Effect of Pyocyanin

The following microorganisms S. epidermis, B. subtilis, Staph. aureus, Micrococcus luteus, Pseudomonas teessidea PR6.5 (AM419154), P. clemancea PR22.1 (AM419155), P. aeruginosa DS10-129 (AM419153), Aspergillus niger and Sacc. cerevisiae were used in this study. The inhibitory effect of pyocyanins was determined by using the agar diffusion method. 60 μL suspensions of each microorganism with an OD of 1 ± 0.2 were used to inoculate 20 ml of sterile molten Antibiotic Medium Number No. 1 Agar (ANTBA1) respectively. The composition of ANTBA1 was (g/l): peptone 6; tryptone 4; yeast extract 3; Lab-lemco powder 1.5; glucose 1; and agar 11.5. The agar plates were allowed to solidify at room temperature and three holes of 7 mm diameter were punched on the solidified inoculated media using a cork borer. 60 μL of the concentrated cell free supernatant harvested at different times was added to the agar wells and incubated at 37°C for 18–24 h. For A. niger and Sacc. Cerevisiae, Sabouraud’s Dextrose Agar (SDA) was used and the plates were incubated at 30 ºC for 18–48 h, respectively. The suspension of B. subtilis was prepared from a culture that had previously been incubated for 5–7 days to allow for the formation of spores. The antimicrobial activities of pyocyanin against each organism were determined by measuring the diameter of the zones of inhibition using a pair of vernier callipers.

2.7 Production and Characterisation of Biosurfactants

The produced biosurfactants were detected by using the following quantitative and qualitative methods.

2.7.1 Quantitative Estimation of Rhamnolipids

The concentration of rhamnolipids in the sample was estimated by the Orcinol method [16][17]. The rhamnolipids concentration was quantified from the standard L-rhamnose calibration curve between 0 and 50µg/ml and the result expressed as rhamnose equivalents (RE) (mg/ml) by multiplying rhamnose values by a coefficient of 3.4, obtained from the correlation of pure rhamnolipids/rhamnose[17][18][19].

2.7.2 Surface Tension Measurements

The surface tensions of the cell free supernatant medium were measured using a Tensiometer (Kruss Digital Tensiometer - Model K9, Germany).

2.7.3 Determination of the Critical Micelle Concentration (CMC)

Critical micelle concentration (CMC) is the lowest concentration of the surfactant that is required to form micelles. When a surface is saturated with molecules of surfactants any more addition will result in the arrangement into micelles and this property is a characteristic feature of biosurfactants. The determination of CMC was performed by diluting the cell free supernatant containing the highest concentration of rhamnolipids and each sample was diluted up to 1: 512. The surface tension of the diluted samples was measured at room temperature (25°C) as described previously using a tensiometer.

2.7.4 Haemolytic Activity

The haemolytic activity of rhamnolipids was estimated by using agar diffusion method. Four equidistant holes of 7mm in diameter were punched on petri-dishes containing 5% Sheep Blood Agar. 60µl of the cell free supernatant extracts was introduced in each well respectively. The inoculated plates were allowed to stand at room temperature for 24-48h. The zones of haemolysis were measured for each system.
2.7.5 **METHYLENE BLUE AGAR ASSAY**

The concentration of anionic surfactants in aqueous solution was determined by the formation of an insoluble ion pair with cationic substances. The insoluble ion precipitates in the Methylene Blue Agar by forming a deep dark blue color against a light blue background using the cationic substance CTAB [20]. Rhamnolipids were identified by formation of dark blue hallow zones on a light blue background by agar diffusion method [19].

2.7.6 **EMULSIFICATION ACTIVITY OF RHAMNOLIPIDS**

The emulsifying activity of the cell free supernatant was determined by using the approach practiced by [21]. 2ml of supernatant and 2ml kerosene mixed and vortexed for 2 min. The mixture was allowed to settle undisturbed at room temperature for 24 hours. The height of the emulsion layer was measured after 24h to determine the emulsion index.

2.7.8 **INFRARED SPECTROSCOPY**

The molecular structure of the rhamnolipids was identified by using a Fourier Transform Infra red spectroscopy (FTIR) Perkin Elmer 100 series. The cell free supernatant was acidified to pH 2 by adding 2M Sulphuric acid to precipitate the rhamnolipids. The precipitated rhamnolipids were extracted with an equal volume of 2:1 dichloromethane/methanol. The organic phase was dried with anhydrous Sodium Sulphate (Na$_2$SO$_4$) to remove excess water and was evaporated to yield the pure rhamnolipids on a rotary evaporator (Buchi, rota vapor R-200 Germany) set at 60-70°C. 2-5mg of the concentrated rhamnolipids was characterized on the Perkin Elmer Spectrum 100series FTIR spectrophotometer.

3. **RESULTS AND DISCUSSION**

The growth profiles for *Pseudomonas aeruginosa* under the three different environments are shown in (Fig 2). The fermentation process in the microbioreactor was limited to 24h while for the conventional scale it was 120h. The progress of the batch fermentation process of DS10-129 in the bioreactor, shake flask and the fermenter was monitored by measuring the optical density of the spent medium. The generation time for *Pseudomonas aeruginosa* in the microbioreactor was slightly higher when compared with the other processes. This development could be attributed to the small cross sectional dimensions of the microbioreactor which results in rapid heat and mass transfer due to the large surface area to volume ratio. Production of pyocyanins was evident during the exponential phase (6 h) of growth of the organism in the bioreactors. Pyocyanin production showed a steady increase in concentration throughout the incubation period (Table 1). In view of this development the early production of pyocyanins could have been caused by an increase in temperature in the channels. Previous reports [22] have shown that the production of pyocyanin is dependent on increase in temperature.

**Table 1 Antimicrobial effects of pyocyanin on microorganisms**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Antimicrobial activity in terms of diam. of zone of inhibition (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 h$^a$</td>
</tr>
<tr>
<td><strong>µBR</strong></td>
<td>SF</td>
</tr>
<tr>
<td>Staph. epidermis</td>
<td>24</td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>18</td>
</tr>
<tr>
<td>Micrococcus luteus</td>
<td>16</td>
</tr>
<tr>
<td>Staph. aureus</td>
<td>13</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>10</td>
</tr>
<tr>
<td>Pseudomonas teessidea</td>
<td>9</td>
</tr>
<tr>
<td>Pseudomonas clavancnea</td>
<td>8</td>
</tr>
</tbody>
</table>

$^a$ Incubation time, µBR = microbioreactor, SF = shake-flask, F = fermenter

Aspergillus niger: initially small zones were observed, but disappeared after 48 h, *Pseudomonas aeruginosa*: No zones of inhibitions were observed.
FIG 2. Growth of *P. aeruginosa* DS10-129 in Glycerol Supplemented Nutrient Broth (GSNB) medium in three bioreaction conditions. The OD₆₀₀ values of fermented culture from each bioreactor were determined at various time intervals at 30°C. The data represented is a mean of three fermentation measurements in each bioreactor.

FIG 3. Inhibition zones of *Bacillus subtilis* by a 24h concentrated cell free supernatant. A = Shake flask (1), B = Fermenter, C = Microbioreactor, D = Shake flask (2)

FIG 4. Surface tension measurements of aliquots of cell free supernatant at various incubation times were determined for each bioreactor. The data represented is a mean of three fermentation measurements in each bioreactor.

FIG 5. Critical Micelle Concentrations of the biosurfactant concentrated from different fermentation trials. Indirect measurement of rhamnolipid concentration by determination of the critical micelle concentration (cmc). Samples of the cell free supernatant were diluted 500 fold and their surface tension measured for each dilution.
FIG 6. Haemolytic activity of concentrated cell free culture supernatant on agar plates containing 5% sheep red blood cells (fermentation was carried out in a batch fermenter). A = 8h, B = 16h, C = 24h, D = 24h, E = 72h and F = 120h.

FIG 7. Estimation of rhamnolipid production by *Pseudomonas aeruginosa* DS10-129 during fermentation carried out in a 100ml Shake Flask. The methylene blue agar was used to confirm the presence of rhamnolipids production by the organism. A = 24h, B = 48h, C = 72h, D = 96h and E = 120h

The amount of biosurfactants produced by DS10-129 was estimated by using the Orcinol method. The maximum production of rhamnolipids is shown in Table 2. The summary result of the rhamnolipids quantities produced by the organism in the bioreactors is consistent with the surface tension measurements.

<table>
<thead>
<tr>
<th>Type of Bioreactor</th>
<th>Cell free Supernatant collected after media</th>
<th>Surface Tension (mN/m)</th>
<th>Emulsification index (E_{24}) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>µBr</td>
<td>16h</td>
<td>27.9</td>
<td>92.1</td>
</tr>
<tr>
<td>Shake Flask</td>
<td>96h</td>
<td>29.6</td>
<td>84.7</td>
</tr>
<tr>
<td>Fermenter</td>
<td>96h</td>
<td>28.3</td>
<td>78.6</td>
</tr>
</tbody>
</table>

Table 2. Quantification of biosurfactants produced by DS10-129.

FIG 8. Infrared Spectra of rhamnolipids: A = 3368.10 cm⁻¹ (OH free stretch due to hydrogen bonding), B = 2925.42-2856.40 cm⁻¹, (Aliphatic bond stretch CH₃ CH₂ and CH), C = 1737.33 cm⁻¹ (C=O stretch due to the ester functional group), D = 1455.52-1380.65 cm⁻¹ (bending of O-H bands in carboxylic acid group), E = 1300 – 1033.61 cm⁻¹ (C-O stretch due to ester functional group).
Results in Table 2 and Figure 4 show that DS10-129 produced effective biosurfactants that lowered the surface tension of GSNB from 58 to 27-30mN/m. The result of the critical micelle concentrations of the diluted samples in Figure 5 shows a reduction of surface tension of distilled water to about 30mN/m, indicating that the tension-active properties of the rhamnolipids produced have better prospects for industrial applications such as bioremediation.

Haemolysis of blood agar was first described [24] to screen the biosurfactants producing bacteria. [25] used the method to quantify the amount of surfactin produced by B. subtilis and demonstrated that the presence of the biosurfactants caused the lyses of the red blood cells in the agar. In this work the presence of biosurfactants in the samples was qualitatively detected by haemolysis of erythrocytes in 5% Sheep blood agar medium. The results show that the cell free supernatant collected from the bioreactors contained fairly large amounts of factors with haemolytic activity as shown by the diameters of zone of clearing on red blood cells in sheep blood agar in Figure 6 for the different sampling times.

Like all synthetic surfactants rhamnolipids consists of a polar head and non polar tail groups, when they combine with different cationic substances, such as Cetyl Trimethyl Ammonium Bromide (CTAB) to form insoluble ion pairs in aqueous solution, which precipitates as dark blue zones against a blue background in the methylene blue agar [21]. The result shown in Figure 7 indicates the production of effective rhamnolipids by DS10-129, which precipitates the CTAB in the Methylene blue agar. Several previous studies have shown that rhamnolipids are normally produced during the stationary phase of growth [4][26][27][25]. The size of the zone of inhibition was proportional to the amount of biosurfactants present in the sample.

Results in Table 2 shows that the biosurfactants produced by DS10-129 has good emulsifying activity. The higher emulsification index corresponds to the complete emulsification of the oil phase. Previous reports [28] have shown that a good bioemulsifier is the one that has an index of 50% and more. The results suggest that the quality of biosurfactants produced by DS10-129 has higher emulsifying activity and would be very useful in bioremediation work. The production of rhamnolipids was confirmed by FTIR spectroscopy. The FTIR spectrum of the rhamnolipids is shown in Figure 8. According to the results of the infrared spectra in Figure 8, the rhamnolipids extracted belongs to the glycolipid group, which is made up of aliphatic acid and ester. The results obtained are consistent with the structure reported by [29] consisting of aliphatic acid and the glycolipid moiety.

4. CONCLUSION

The results obtained from the work have demonstrated that Pseudomonas aeruginosa DS10-129 is capable of producing extracellular secondary metabolites such pyocyanins and biosurfactants when cultured in a microbioreactor environment.

ACKNOWLEDGEMENTS

Authors wish to thank the University of Teesside sponsored Research and Enterprise Development Fund and Higher Education Innovation Fund (HEIF) for their support towards the completion of this project. PKSMR wish to thank UK- Bioscience for Business KTN for the award of FROPTOP fund to further explore the biocatalytic study of biosurfactant production from renewable resources.

REFERENCES


CONTACT AREA, PRESSURE DISTRIBUTION, AND MECHANICAL STABILITY IN EXTERNAL ARTHRODESIS OF THE ANKLE JOINT

C.J. Connor,1* S. Hodgson2 and F. Nabhani2

1School of Computing, Engineering & Information Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
2School of Science & Technology, University of Teesside, Middlesbrough, TS1 3BA, UK

ABSTRACT

The ankle joint is often affected by arthritis, giving a joint that is painful, stiff, and restricts movement. This can result in a huge loss of mobility for the sufferer. Unlike replacement of the hip, the replacement of a diseased ankle joint is not as straightforward and the outcomes do not reach the same success levels. The preferred surgical choice is arthrodesis, a procedure whereby the two bones forming the joint are fused together to eliminate the joint and hence pain. The success of the procedure is dependent upon several factors, two of the most significant being the levels of contact area and pressure achieved during the compression period, during which bone growth occurs across the two bones being compressed together. These factors influence joint stability and micromotion at the bone to bone interface during this growth phase.

This study investigates the levels of contact areas and pressures that can be achieved for different arthrodesis variables. These variables include the joint shape, which can be curved or flat, and the position of the compression pin within the talus, namely anteriorly or centrally positioned with reference to the talar dome. Influence of the Achilles tendon in joint stability is also investigated.

A test rig was developed allowing load/deflection curves to be determined for various configurations of these variables. Models representing the bones under consideration, together with pressure sensitive film, allowed measurement of contact areas and pressures within the joint under compression, achieved using pins and instrumented compression rods.

Results indicate there is little significant variation in contact area and pressure for the different shaped joint cuts, however, if the talar pin is placed in a more anterior position then the contact area can be improved over a centrally positioned pin. Anterior pin placement also gives increased resistance to motion and mechanical stability. It has been established that the arthrodesis construct is especially weak in terms of rotation about the tibial axis, and the results from this study indicate that through the use of a curved joint shape the resistance to this motion can be improved greatly.

1. INTRODUCTION

The ankle is a joint that is often adversely affected by arthritis, resulting in a painful joint with limited mobility. One of the standard procedures for alleviating this pain is arthrodesis of the ankle, which involves fusing together the two major bones that form the ankle joint.

A number of techniques are available in order to perform this procedure, the first recorded technique having been documented by Albert in 1882 [1]. These various techniques include the use of internal screw methods and external compression using pins. Whilst these techniques have been documented and successfully performed for many years, there has been little investigation into the biomechanical factors that determine their success rate and their potential effect upon the joint during the initial bone growth phase.

This study investigated the effect of two variables upon the external arthrodesis structure. The particular fusion technique considered obtained compression at the joint site by inserting one horizontal pin through the talus and another horizontal through the distal tibia, and using compression clamps to force the two pins, and hence the joint, together. In 1951 Sir John Charnley proposed that optimum ankle fusion could be attained by cutting opposing...
bones in a flat shape, and that by positioning the talar pin in an anterior position, the Achilles tendon force would result in an even anterior-posterior compression force distribution [2]. Recent variations on this theme have sought to address the weakness of such a technique, especially when subject to axial rotation loads, by cutting the bones in a curved shape (i.e. attempting to retain approximate original joint shape) [3,4]. This study investigated the effect of pin position and joint shape upon the pressure distribution attained within the joint and its resistance to motion, both factors being critical to successful fusion, as any motion at the joint site will damage any potential bone growth taking place across it.

The vast majority of biomechanical studies in this field have compared the success of various arthrodesis techniques against one another [5,6], with only more recent studies investigating the effects of different physical variables available during the procedure [4]. This study used a combination of internal joint pressure measurement, together with bending and torque deflections at the joint site to determine the likely effect upon joint stability of these variables. The use of finite element models to evaluate the stability of arthrodesis constructs has only recently been applied to internal compression arthrodesis techniques only [7].

The use of pressure sensitive film to determine pressure levels within joints has been well documented in relation to biomechanical studies [8], and this technique was used to determine pressure levels within the joint under compression.

2. MATERIALS AND METHODS

A 3D CAD system was used to develop models of the talus and distal tibia to be constructed from CT scan data obtained of an adult male. These models were resected within the CAD environment to generate suitable flat and curved shaped models of both bones using appropriate surgical advice. Positions of compression pin holes were also mapped on to these computer models. Data from these models was then output and a CNC machining system used to reproduce approximations of these models as physical entities, as illustrated in Figure 1.

![Physical Models](image)

A test rig was developed that would allow the simulation of external arthrodesis. The talar model could be fitted into a component representing the foot, and the horizontal talar compression pin inserted through it. The model representing the distal tibia was mounted in a plate above this in the frame of the test rig. The contact between the lower tibial surface and the superior talus surface allowed the formation of the ankle joint to be simulated. Instrumented compression clamps were then used to apply known compressive loads to the joint, by compressing the talar model against the fixed tibial surface. Compression could also be applied at the rear of the foot component via an instrumented compression rod between the rear of the foot component and the upper frame of the test rig.
Pressure sensitive film (a film that changes colour in relation to applied pressure) was inserted into the joint between the two models and the joint compressed to predetermined levels in order to record pressure distribution within the joint. This procedure was first performed for a flat joint shape with compression being achieved using an anteriorly positioned talar pin. The pressure sensitive film was then removed from the joint in its expended state for evaluation. The flat shaped model was then used again, with a centrally mounted talar pin. The process was then repeated using the curved shaped models, with the pin again being positioned in both the anterior and central positions. These studies were performed firstly without any Achilles tendon compression, and then repeated with the application of an Achilles tendon compressive load.

The pressure sensitive film that had been exerted to compression was scanned, converted to greyscale values and computer programs used that would count the number of pixels and their corresponding greyscale value in order to determine joint contact area and average pressure. Known calibration charts were used to quantify these figures. An example of this output is displayed in Figure 3.

The mechanical stability of these arthrodesis constructs was then measured by compressing the joint as previously, but without pressure sensitive film in the joint. External loads were then applied in order to induce motion at the ankle joint in the various directions that the actual joint can move in, namely
plantarflexion/dorsiflexion, inversion/eversion, and internal/external axial rotation. The level of displacement in such directions was measured in response to increasing loads, allowing load/deflection data to be obtained for the various directions.

3. RESULTS

3.1 PRESSURE DISTRIBUTION

It was discovered that only small differences exist in the levels and distributions of pressure within the joint far varying pin positions and joint shape. Following a MANOVA analysis (p=0.05), it could be stated that:

When the opposing bone surfaces are cut with flat shapes, the optimum contact pressure was obtained with an anterior pin and an applied Achilles tendon load, and that in all cases the Achilles tendon load increases the joint pressure.

When the curved joint shape is retained, optimum contact pressure was obtained with an anterior pin. Again, the Achilles tendon load increases pressure in all cases when a curved shape surface is maintained.

Overall it can be stated that for curved cut models, pressures remain in the main constant with slight improvements possible when using anterior pins. For flat models, overall pressures are slightly higher than for curved joint shapes, and increase even more when using anterior pins.

These results are illustrated in Figure 4.

![Figure 4: Pressure Distribution Results](image)

3.2. MECHANICAL STABILITY

Figures 5 to 7 show a summary of the results of maximum deflection obtained for the various stability tests following one-way ANOVA analysis (p p=0.05). The matching opposite motions i.e. dorsiflexion against plantarflexion were found to follow a similar pattern to the data shown and so only results for one of the motions are provided.
The mechanical stability tests demonstrated that the joint deflection for directions in which physical opposition to motion exist due to the physical formation of the joint, such as plantarflexion/dorsiflexion and inversion/eversion, the deflections recorded were very small (below 0.3mm). These did however illustrate that the Achilles tendon was important in resisting motion and balancing the joint. These tests showed that the weakest combination of variables occurred when a central pin was combined with a flat joint shape, the optimum scenario being an anterior pin and curved joint shape. In between these two extremes came the other combinations i.e. flat with anterior pin and curved joint with a central pin.

The most significant findings however came in the rotational tests where it was clearly demonstrated that a curved joint shape had a major effect upon resistance to motion, which was noticeably larger than in the other loading directions. When coupled with an anterior pin noticeable improvements in resistance to motion were found. The worst situation was again a flat cut with a central pin. The deflections recorded during the rotational tests were in the order of magnitude almost 10 times larger than those measured during the inversion/eversion and plantarflexion/dorsiflexion tests, illustrating the importance of this load factor in resistance to joint micro-motion, and the weakness of such constructs in relation to this type of motion.
4. DISCUSSION

A number of biomechanical studies have sought to define the optimum technique for arthrodesis. These studies have succeeded only in confirming that specific pin, screw or plate configurations performed better than one another. Miller [6] investigated the effect of joint shape, but could not find any statistical differences between flat cut and contour preserved specimens, possibly due to the low number of cadaveric test specimens available. The lack of availability of consistent cadaveric specimens led to this study using polyurethane models to represent the bone. Further validation of this experimental work is currently being performed using finite element modelling, and the results from these experiments correlate well with those found in other finite element studies [7].

This study has also considered a situation of perfect joint congruency, with solid bone material that is not affected by degeneration, however in most rheumatoid cases requiring surgery the joint surfaces remain intact within a painful joint due to cartilage destruction. Investigation into the application of different material properties representing poor bone stock would be a useful addition to this work.

5. CONCLUSION

It can be seen from Figure 4 that the pressure sensitive film tests showed only slight variations in pressure across the joint for different pin and joint shape combinations. They did however illustrate the importance of considering the Achilles tendon during such experiments. It was discovered that marginally better results were obtained using an anterior pin position, with little difference between curved and flat joint shapes.

The mechanical stability tests however clearly illustrated the benefits that could be obtained in resistance to motion across the joint when using an anterior pin and curved joint shapes. This was especially true for rotational resistance, which is the major weakness of arthrodesis constructs. Whilst there may be a slight reduction in joint contact pressure when curved models are used, these differences are insignificant in comparison to the additional joint stability that can be afforded in rotational displacement when such surface shapes are retained.

These findings have been further substantiated by additional 2-dimensional finite element studies that have been performed, and confirm the hypothesis that a curved shape and an anterior pin will give greater stability during the arthrodesis recovery period, improving likelihood of fusion.

Ongoing work includes the development and utilisation of 3-dimensional finite element models of these joints in order to further study the contact across this joint both in terms of pressure and displacement under load during compression.

REFERENCES

ABSTRACT:
Whiplash injured patients often complain of pain in the shoulders following rear impact, it is the aim of this investigation to assess the cause and effect of these pains and advise on corrective measures. A 3D model was generated and a finite element analysis was performed simulating a crash situation to assess and define this injury more clearly. The model displayed a clear proximal force resulting from the impact scenario which created an increased loading on the supraspinatus. This injury occurs when a driver sees a vehicle approaching from behind and braces himself/herself for the impact locking the elbows and pressing against the steering wheel. This causes the shoulder centre to become posteriorly and proximally located so that during distal impact, large forces are impacted on the supraspinatus and subscapularis tendons. It is hoped that with the information gathered from this study that these injury's will be diagnosed earlier and lead to better long term recovery.

1. INTRODUCTION:
The number of rear impact crashes in the UK is alarming given the relatively low protection drivers are offered in current Automobiles. According to A. P’erez del Palomar in 2007 Rear-end impacts account for more than one-third of vehicle Accidents [1]. Simpson (1994) found that 16.1% of road crashes occurred from a rearward direction. This figure may be conservative since inclusion in this study required a minimum claim of £202 [2]. Rear end impacts are most common at intersections where the struck vehicle is stationary or travelling slowly for a red light [2].

The normal recommended driver seating position is displayed in fig 1. This position maximizes control and safety while driving. The braced position as seen in fig 2 is when the hands are placed firmly against the steering wheel, elbows locked and the body pressed firmly back into the seat back. Bracing often also forces the body upwards in the seat to allow for full extension of the arms to the locked position, this action pushes the shoulders above the support of the rear of the seat. This is a natural reaction when the body is in danger.

Though the technology has been available since 1999 very few cars are fitted with rear impact sensors. During a frontal impact frontal sensor activate the deployment of air bags. This not only absorbs impact but is designed to force the hands off the wheel or at least bend the arms. This stops people locking elbows during impact.

The injury in question in this paper is a pain in the shoulder immediately after impact when hit from the rear in the braced position. The injury is often diagnosed as a nervous condition in accident and emergency or is treated merely with anti-inflammatory medication. This injury has symptoms similar to bracing injury’s occurring from referred pain from the neck; shoulder symptoms often arise from both sources. However there are cases reported by Frank [3] during clinical experience of people suffering pain and immobility 2-4 years after the injury occurs.

* Corresponding author: Tel.: (01642) 342482; E-mail: f.nabhani@tees.ac.uk
It is proposed that when a rear impact occurs, while in the braced position with the combination of the biceps, triceps and deltoid restricting forward joint motion and the glenoid restricting backward motion of the proximal humerus that the shoulder centre becomes posteriorly and proximally located putting large crushing forces on the supraspinatus and subscapularis tendons muscles. These muscles then scar which makes them more susceptible to tearing which leads to a greater build up of scar tissue in the joint. This supposition would seem concurrent with the cases detailed above where problems only arise 2-4 years after trauma [4]. The Stretching Institute [4] advise patients with muscle injuries that to prevent the buildup of scar tissue the injured area needs to be treated within the first 72 hours. If scar tissue develops it creates a weak spot in the muscle and loss of flexibility [4]. These are typical symptoms in cases of this rear impact shoulder injury as observed by Mr P L Frank.

It is reported by Frank [4] during his clinical experience that people who suffer from this injury tend to be males. This is probably attributed to males sitting with the seat further back and often with the seat more reclined making the locking of the elbows easier. This observation is supported by the findings of Parkin et al [5] who measured driver seating position on 1000 cars and noted that men tend to sit further back in their seats.

Given the high prevalence of rear impact crashes and the clear force applied to the upper limbs during impact it is important that drivers are properly protected during a crash situation. Euro NCAP (European New Car Assessment Program) however does not even consider shoulders in their testing “Neither arm carries any instrumentation in a crash test, the arms flail around in an uncontrolled way” [6].

It is often assumed that occupants with advance warning of an impending collision have fewer and less severe symptoms than those who are surprised by rear-end impacts [7]. However this currently unrecognized injury could change those statistics.

Aims:
1.) The cause of pain generated in the shoulder complex during reverse impact road accidents
2.) Loadings generated on the supraspinatus and subscapularis tendons.

2. METHODOLOGY:

2.1 EXPERIMENTAL MODAL

In this study a 3-Dimentional Finite Element Analysis (FEA) model will be generated to recreate the in-vivo mechanics of the glenohumeral joint during the impact phase of a rear impact automotive crash. FE models need a great deal of computational power, but can provide detailed information about tissue deformations and injury prediction. Multi-body models can also include many anatomical details while being computationally efficient. This makes them suitable for parameter variation and optimization analyses [8]. The value of FEA in modern prosthetic design and clinical testing cannot be understated. 3D in-vivo Models are accurate and reliable [9] providing opportunities for complex testing to be performed in almost any theoretical situation.

2.2 LOADING VALUES

The loadings to be applied to the Model are taken from previously generated research, this provides a accurate peer reviewed set of constraints for testing.

The moment generated by the contorting body is based upon findings by Golinski et al (2004) who proposed a moment of maximum value 1.5Nm based on sled experiments conducted by Steffan [10-11]. Measurements from real accidents were used in order to achieve realistic deceleration of the sled.

The model also uses the acceleration pulse suggested by Euro NCAP of 4.4 m/s applied to the seat [12]. This is the force generated as the seat is catapulted toward the steering wheel by the impact.

The rotor cuff will be treated as a constant 15N force across the joint [13]. This constant, evenly applied force is typical of tests of this type balancing the shoulder and recreating a basic loading similar to that provided within the joint.

It has been observed that Muscle tensing can change the effective stiffness and mass of a body segment or region [14]. A 338 per cent increase in thoracic stiffness (from 70 N/cm to 236 N/cm) was observed when the volunteers maximally tensed the muscles of their shoulders, thorax, arms, back, and neck [14]. In this light the shoulder joint will be treated as fixed restrained in the scapula-thoracic plane as any movement is heavily restrained by the surrounding musculature and the seat back.
It has been shown that the long head of the biceps and the deltoid muscle have an important role in the stability and support of the glonohumeral joint [15-17]. This is represented in the model using a 700kPa distributed force over the muscle contact areas. 700kPa is an approximated muscle force value used by R W Kent et al in their 3D FEA model for testing thoracic deflection described above [14].

### 2.3 Car Seat Model

An accurate modeling of the car seat is another essential aspect of rear impact investigation [12]. Car seats comprise of a backrest, a sprung centre and a foam cover. The material properties defined for the chair back were based upon data collected by Nicolas Bourdet et al (2007) in their design for a car seat model for rear end impact testing [12] See figure 3.

<table>
<thead>
<tr>
<th>Young’s modulus</th>
<th>230MPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson’s ratio</td>
<td>0.29</td>
</tr>
<tr>
<td>Density</td>
<td>7800 kg/m3</td>
</tr>
</tbody>
</table>

Figure 3. Table showing car seat back properties.

Golinski et al stated that only once the limit of seat back deformation is reached does the torso rapidly accelerate forward [11]. However due to the nature of the bracing position the torso is already forced firmly back into the seat to allow the elbows to lock fully (fig 2). For this reason it was determined to neglect the effects of cushioning in the seat during impact in this study as its presence is deemed negligible. The seat in the generated model forms a solid fixture behind the scapula which restrains the impact forces.

### 2.4 Shoulder Position

The dimensions of the shoulder complex were taken from CT scans produced for the Laboratory of Human Anatomy and Embryology, University of Brussels. These were used as they provide a high quality, previously tested, set of images.

There are no specific angles at which the arm can be said to be at as it depends on driver height, driving position and seat recline angle. The theoretical driver was selected to be a 50th Percentile Male Driver. As stated earlier this injury is most often detected in male drivers. The theoretical driver arrangement for this study can be seen in fig 4.

For the purposes of this study the humerus was said to be at an angle of 15º to the scapulothoracic plane shown as Y in fig 4.

Figure 4. Assumed Driver Seating position.

### 2.5 Shoulder Model

A threshold was applied to the CT images (1 slice per mm) and polylines generated around the model extremities 1 per slice.

The Shoulder complex was then modeled in IDEAS-nx12 (Siemens PLM). This allowed for modeling and finite element analysis to be performed in a single software package. The polylines generated were lofted together to create a solid model.
The model was Free meshed with 2.15983 mm Solid tetrahedral elements as shown in figure 5. Trabecular bone has not been accounted for in this study.

![Meshed Glenohumeral Joint](image)

Figure 5. Meshed Glenohumeral Joint

### 2.6 Testing

The model was loaded according to the in vivo force mechanics discussed in 2.2. The back of the seat and the base of the seat were treated as fixed. An image of the loaded model is displayed in figure 6. The FE model was then solved using the IDEAS_nx 12 FEA software which generated no clear errors in the model or corrupt elements. Corrupt elements are the main contributor to incorrect results in FEA and in a complex shape like the proximal humerus it is particularly important that element checks are performed.

![Assembled FE Test showing loadings and restraints.](image)

Figure 6. Assembled FE Test showing loadings and restraints.

### 3. RESULTS:

The results generated by the IDEAS_nx 12 software are presented in vector components as displayed in figure 7. This information is also displayed in figure 8 which shows the displaced humeral head in the light grey area.

<table>
<thead>
<tr>
<th>Max Displacement:</th>
<th>X</th>
<th>Proximal</th>
<th>1.73E-1mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y</td>
<td>Lateral</td>
<td>1.27E-1mm</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>Posterior</td>
<td>2.25E-1mm</td>
</tr>
</tbody>
</table>

Figure 7. Displacement Results for FE test rig.
Bracing Shoulder Injuries caused by Rear impacts.

4. VALIDATION OF MODEL

This model will be validated in two ways, firstly by model comparison and secondarily by a mechanical test rig. In a paper by Bryce et al [9] a validation of a scapula model was attempted. Using very similar methods of model generation, a model from CT scans was generated. All results were within 3mm accuracy from the original sample. They concluded that “clinically accurate 3D images can be created from CT scans” [9] and that “the results from this work verify that accurate and reliable 3D models can be created from in situ scapulae” [9]. A mechanical test rig is currently being designed which will recreate the moments and forces encountered during this form of rear impact to fully validate the results.

5. TEST RIG DESIGN

The mechanical test rig must meet certain criteria based upon current test rigs and empirical data to accurately reproduce the in-vivo mechanics of the glenohumeral joint. The test rig must accurately recreate the loadings generated during the rear impact crash scenario applying the relevant loadings representative of the in vivo physiological characteristics of the Glenohumeral joint. The test rig must also simulate the torsional loading / deforming forces applied to the proximal humerus generated when bracing to recreate the dynamic muscular stabilization of the shoulder girdle along the line of action of each of the muscles. The test rig that is to be built is a modular, dynamic design allowing for the testing of multiple activities of daily living and directly applied stresses on the glenohumeral joint and potentially other joint in the body. Figure 9 shows a basic structure for the test rig, an explanation of the different part functions is displayed below.
Mechanical Test Rig Design Breakdown.

<table>
<thead>
<tr>
<th></th>
<th>Rotational Disk</th>
<th>Glenoid Support Column</th>
<th>Glenoid Holder</th>
<th>Proximal Humerus Mount</th>
<th>Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The disk is a slim lined rotating table, mounted on the side of the test rig. It allows for accurate and strong rotation of the glenoid support column.</td>
<td>The glenoid support column supports the glenoid holder. This column also allows for angular adjustment of the glenoid holder. This enables the glenoid holder to face any direction. The column locks to form a solid support when testing.</td>
<td>The glenoid holder supports an accurate molding of the glenoid which is removable for use in testing. The holder also has an array of attachments to which much of the key musculature around the glenohumeral joint can be attached during testing.</td>
<td>The Proximal Humerus mount simply supports and locks in a model of the proximal humerus for use in testing. Like the glenoid holder it also supports attachments to support musculature and uniquely for the rear impact test it supports a small chair back model.</td>
<td>The Module is interchangeable with each test. Each module has a different purpose for example to re-created the action of an activity of daily living of the direct, blunt impact force of a fall. In the test for the rear impact injury a pneumatic cylinder is firmly mounted to the base of the test rig. This cylinder is computer controlled to recreated the force and acceleration of the impact.</td>
</tr>
</tbody>
</table>

During the testing of the rear impact crash while in the braced position the model will recreate the moment forces from the impact, it will aim to recreate the tensed state of the muscles and the effects of the support back. The forces are to be recreated using a computer controlled pneumatic cylinder module. The musculature is to comprise of thin steel cables surrounded by a dense polymer cover tensed between the relevant test rig fixture and the muscle insertion point. Muscle tension will be measured using torque wrench when tensioning the cables. The seat back is simply a Glass filled Nylon block covered in a 1mm layer of 100% polyester fabric. This is aimed to look at the forces the seat back has on the displacement and forces in the shoulder and so is merely representative of a seat back. As discussed in 2.3 exact seat model properties are not required for this test. In order to ensure reliable, accurate results the mechanical test rig will be validated against 3D motion, loading and deformation data from previous literature.

6. Discussion:

It was the aim of this investigation to try to identify the cause of pain generated in the shoulder complex during reverse impact road accidents.

This is a previously unexplored injury and hence we hypothesized that when braced the shoulder centre becomes posteriorly and proximally located so that during distal impact, large forces are forced on the supraspinatus and subscapularis tendons. The rationale for this hypothesis is that the combination of the biceps, triceps and deltoid restricting forward joint motion and the glenoid restricting backward motion of the proximal humerus the head of the humerus to become posteriorly and proximally located under the forces generated in a automotive rear impact.

A similar study was performed by Scarlat et al. looking at the effect of lateral impaction to the shoulder. FEA was used with retrospective case studies and discovered that when “associated lesions” are considered better outcomes are generated. The follow up conducted however only took place 6 months after trauma indicating that possibly the muscular crushing indicated in this paper was not considered or diagnosed though one patient had developed AC arthrosis [14].

To prove the concept of the crushing injury a finite element analysis was carried out restricting motion using the glenoid and the surrounding muscles and balancing the shoulder with a rotor cuff constant 15N force [13]. The Model displayed a maximum 1.73E-1mm proximal and 2.25E-1mm posterior displacement of the humeral head shown in figure 3.1. Though these results prove the hypothesis it is debatable whether such a displacement is significant enough to crush the supraspinatus and subscapularis tendons.

The model is limited by the fact that trabecular bone has not been accounted for in this study. This was chosen to make the model possible to solve using the current FEA package. Future studies should look at the effects of
Bracing Shoulder Injuries caused by Rear impacts.

...trabecular bone on the movement of the proximal humerus. The model also only includes the glenoid rather than the whole scapula, this resulted in the ignoring of any transferred stress into the scapula and its surrounding muscles. As discussed in 2.2 it has been suggested that the scapula can be considered fixed when muscles are tensed when anticipating a strong impact [14]. This omission is unlikely to affect our results in this light.

The complex musculature around the proximal humerus made the accurate application of force difficult to balance the joint. The use of an evenly deployed rotor cuff load simplified the model greatly and is generally accepted practice [13].

The ability to directly and accurately import the CT images is key to this project. The import of CT data produced a very accurate model of both the proximal humerus and the glenoid. FE models of this accuracy and quality are comparatively new to the bio-mechanics field because until recently the import of data of this format has been lacking in accuracy and ease of use. The ability to easily import CT data to CAD packages has dramatically impacted the bio-mechanics field making it possible to quickly render CT and MRI data for testing. The IDEAS software was also key to the success of the FE model as its high processing power allowed for very fine element sizes across the complex shape.

The model is a very accurate representation of the in-vivo forces generated during an impact of this type and further work on the model should allow for an expansion into an accurate, dynamic shoulder model. The model should also begin to account for associated injuries from lateral impaction forces which depend on the direction of the initial force, the magnitude of the force and on the compliance of the soft tissue at the point of impact [18].

The mechanical test rig when built will form a platform to validate the FE model under any load or motion. The simple design reduces the possibilities for user error and allows it to be a widely usable tool. The test rig is a portable size meaning that it will be possible to use it not just for research but for the teaching of medical and bio-mechanical students.

6. CONCLUSION:

As a result of the study the crushing injury proposed has been shown to be a real possibility. Given that such a high percentage of crashes in the UK are rear impact [1] it is important that it is properly diagnosed. This study recommends that the injury should be checked for in Accident and Emergency when patients come in having been in an accident of this type and complain of shoulder pain. Early diagnosis and treatment will avoid complications later in life. The mechanical test rig will provide further detail and validation of this injury. Due to the complex nature of the injury and the easy confusion with referred neck pain future clinical investigation will be required to define the mechanism of injury and to confirm that the neck is not responsible for referred pain. It is important to define this injury as it is likely to become more common as shoulder and neck surgeons no longer – or rarely – deal with both areas [19].

It has been shown that the anatomical lesions generated following a rear impact may be misdiagnosed [18]. This study recommends that all rear impact patients should be asked about shoulder pain and that any pain should be investigated.

ACKNOWLEDGEMENTS:

The authors wish to acknowledge Prof R J Minns, for the images of the braced position and Mr. A Rangan, for the clinical support he has offered. We are also indebted to Mr. P L Frank for details for the injury and the clinical observations he has offered. We also wish to thank the Laboratory of Human Anatomy and Embryology, University of Brussels (ULB), Belgium for the image datasets used in this experiment.

REFERENCES:


* Corresponding author: Tel.: (01642) 342482; E-mail: f.nabhani@tees.ac.uk


[19] Private communication – Dr Peter Frank – May 2009
Inventory System Improvement in ERP Implementation Projects

Michael Boyce*, Dr. Nasreddin Dhafr & Robert Lee
Teesside Manufacturing Centre (TMC)
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

Inventory control is one of the most vital areas to manage within a manufacturing company. Implementing an ERP (Enterprise Resource Planning) system gives the organisation the ideal opportunity to further explore and improve this area encompassing stores, purchasing and manufacturing departments.

Some manufacturing organisations nowadays implement ERP and link it to lean manufacturing concepts like JIT (Just In Time), Kanban system and 5S.

The objective of this paper is to investigate how some lean concepts were utilised to assist the implementation of the ERP in the early stages. This was done in a large, multi-national manufacturing organisation. The organisation supplies the oil and gas industry in a niche market.

This work involves the improvements that were made in the warehouse. The main part of this work was done in the Tool Crib. This is a part of the warehouse where all the tools and consumables are stored. These are items that are used in the production process but are not part of the end product. Over the years more and more items have been stored in this area, until now when there is over 5000 items. The work includes the investigation of the rationale for keeping, or not, these parts as well as data collection. The aim been to reduce the inventory by disposing of obsolete items and improve efficiency of the warehouse as a result. The aim of the data collection was to gather data relevant to each of the items. This encompasses a wide range of data stretching from usage history to vendors and their corresponding part numbers.

The work was done and many challenges were met on the way most of these were overcome whilst some are ongoing. This was to be expected as usually in the case of ERP implementation it is very rare to accomplish immediate improvements in all areas.

1. INTRODUCTION AND PROJECT BACKGROUND

Worldwide business competitiveness and the increased speed of information generalisation mean that the integration of enterprise resources and information is necessary for enterprises to survive. Thus, enterprises are exploring the possibilities of integrating their sales, manufacturing, inventory, financial accounting, and human resources into a single managing system [1]. ERP systems assist organisations to integrate and automate these areas [2]. One of the main aims of ERP implementation is not only to integrate the existing business processes into the ERP system, but also to improve them and then implement. As just implementation of an ERP system is no guarantee of optimisation of the company functions. The ERP implementation must be combined with a re-design and re-organisation of the existing processes [3]. Early thinking on Lean often dismissed ERP and other information systems as inconsequential or even inappropriate in a Lean focused company [4]. More recently, as ERP systems have developed, ERP vendors are now claiming it is possible to combine lean principles with the implementation of an ERP system. Bridging the seemingly insurmountable gap between enterprise resource planning (ERP) systems and lean manufacturing [5].

This paper will focus on the first phase of ERP implementation that was carried out in a medium sized company, located in the North East of England. The company manufactures technologically advanced flexible piping for the oil and gas industry and has expanded at a significant rate over the last several years. It was

* Corresponding Author: Tel.: +44 (191) 2959000; Fax: +44 (191) 2959001; E-mail: m.boyce@tees.ac.uk
realised that as the company expanded the need for more efficient Information Technology systems and department integration was essential to maintain progress.

After further investigation of ERP systems it became apparent that outside consultancy help was required who had extensive experience in this area. To this end TMC † (Teesside Manufacturing Centre) was recruited. With their expertise a specific ERP system was chosen which was engineering and project focussed. This was completed by evaluating several ERP systems and then creating a shortlist. The companies on the short list were then invited to demonstrate their wares.

TMC also created detailed process maps of all departments. This was done by conducting interviews with managers and staff from the company. These maps were then analysed and a first phase ERP implementation scoping exercise was carried out. From this it was decided that Engineering, Warehousing, Purchasing and Finance would all be part of the initial implementation drive for first phase. This paper will focus on warehousing and purchasing.

A full review of the methodology that is used by TMC when helping companies with ERP implementation is published in [6].

The scope of this paper will focus on warehousing and purchasing. The biggest issue facing these departments was the organisational culture and the dislike to change.

2. WAREHOUSING

The stores in the company are split into three areas; they are: (1) raw materials, which consists of standard type material, (2) project stock, which is mostly project specific, and includes the end fittings that are of a unique design for each project and (3) the tool crib, which stores all the consumables and tools used in the manufacture of the pipe, these parts do not become part of the pipe.

The first part that was focussed on was the raw material. This consists of a large warehouse with lots of racking holding material such as thick steel wire coils, stainless steel coils and various types of polymers housed in large hoppers. To give an idea of scale, each of the thick steel wire coils weigh nearly 1.5 tonne and each hopper can hold up to 20 tonne of granular polymer. At any one time this part of stores will house in excess of 500 wire coils.

What became immediately apparent was the lack of knowledge of where anything was. There was a general rule that the coils of different types of steel were kept in their own area but there was no record of what stock came in last or what type or weight it was. The only way to find out was to go to the material and read the label. This in itself could be problematic as they were stored up to 10 metres high. This problem was only going to get worse as an expansion was going on at the time and the raw material area was to be more than doubled in capacity. It was obvious that something had to be done immediately.

A benefit of the ERP system is that raw materials can be tracked throughout the warehouse, using unique location codes. Here you are able to input unique code numbers and later able to assign stock items to these locations. Using this system it is possible to know where every part is within the warehouse. To do this a unique location code numbering system had to be implemented. This was first done by numbering each of the storage locations in a rack using a combination of letters and numbers. This was done for every rack and then each of the racks was given a unique identification. This is shown in Figure 1 below.

† TMC is based in the School of Science and Technology at the University of Teesside. TMC specialises in enterprise integration and provides local organisations a service in design, manufacturing and business process improvement.
Each rack was then given an identifying letter so people knew, using a rack map, where it was in the warehouse. It was at this early stage that it was decided to give all the parts in the warehouse a new part number. It was decided to have raw material part numbers start with a “R”, project specific stock to start with a “P” and tool crib stock to start with a “T”. In this way the first identifying letter for each rack was given the same letter to identify which part of the warehouse it belonged to.

Once it was determined what number each location was to be given labels were printed out to stick on the racks. These locations were inputted into the ERP system and so it was possible to print these out through the new system. Doing this meant we could also have a bar code on the label which could become beneficial later.

**RAW MATERIAL**: Before the first phase implementation of the ERP it would take over an hour to book a raw material into stock. Whenever any material arrived two excel spreadsheets and a database first had to be updated. The database would hold information of what was expected to arrive and so this had to be updated. There was then a booking schedule spreadsheet that also had to be updated. There was then the stock control spreadsheet to be updated to keep a record of what was in the inventory. The old process is shown in Figure 2 below.

In addition to these tasks other labels and documentation had to be produced to help identify the material at a later date and also so that the Quality Control department knew what had arrived. This was called the quality
control package and included labelling up the new material. Each batch and type of material in the delivery had to have an individual hand written form filled out. Also hand written tags had to be filled out and stuck onto the material. The delivery note also had to be matched with the purchase order which part of had to be filled in. This process is shown in Figure 3.

![Figure 3: Old style quality control package](image)

To summarise both these processes had to be done for any raw material that was delivered to site. This could mean 50 pallets of material could be delivered at any one time and it was estimated that an equivalent of two stores man were going through the process of receiving raw material 24 hours a day. This was seen as a good opportunity for the ERP system to make large benefits. After the first phase go-live the process for receiving raw material was made much simpler as shown in Figure 4.

![Figure 4: Receiving material after Go-Live](image)

As can be seen, no longer did two spreadsheets and a database have to be maintained. Everything was all in one system. In addition to this benefit it was also possible to know in advance what was to be delivered and when. This helped to manage resources.

Using the ERP system to do the quality control package was even more beneficial than for receiving raw materials. It is now possible to print out an arrival report to match with the delivery note and then to print out as many labels as that is required in seconds, before they had to be hand written. On these labels all the relevant details, including bar codes, were included. The new process for producing the quality control package is shown in Figure 5.

![Figure 5: Quality control package after Go-Live](image)

**PROJECT STOCK:** The project specific stock is stored in its own part of the warehouse but it still had the same shortcomings as the raw materials. The unique location number was also included in this area. In some ways it was more important that it was because of the way the stock was held. The project stock is a mixture of unique and standard items. They are purchased for specific pipes and stored on the same pallet for that pipe. A problem that has occurred in the past would be if an operator could not find all the parts that he required for a specific job then he would take them from another pallet, and so another project. This problem would then snowball, as the same problem would then occur when the parts from the other pallet were required. Using the new ERP system everything is in its own location and to use a part it must be issued from that location. In fact the ERP implementation highlighted this problem and so now there are stricter rules on personnel access to the stores area.
A large problem did occur in this area after the first phase Go-Live. This was that each of these project parts was ordered for an exact pipe, even though usually they are identical if used on the same project. This was done because it is paramount that it is known exactly what pipe a specific part is fitted to for material traceability and so they all have a serial number. If a pipe was moved into the area to fit the project specific parts to and then it was found that it was the wrong one for the part it would get fitted anyway, and so loses material traceability. This issue is covered in more detail in another paper reference. [7]

**CONSUMABLES:** The tool crib is an area that holds the tools and consumables that are used in the manufacture of the pipe but do not become part of the finished product. At the start of the first phase there were just over 5000 separate types of item held in this area and this number was increasing by approximately 30 per month. A small amount of investigation was carried out to determine what items were actually held there, as no one seemed to know. It became apparent that some items had been held in the tool crib for 10 years and never used. It was decided to do a “data cleanse” in the area to capture information on all the parts and then to come up with a list of items to recommend to the company to dispose of. It was deemed that this would also help to determine a finalised list of the items that were required in the new ERP system.

The aims of doing the data cleanse in the tool crib were not only to compile a list of inventory items for the new system but to also help the area to work more efficiently.

The anticipated benefits were:

- To minimise the number of inventory items.
- To minimise the space used.
- To increase the stock efficiency.
- To increase the accessibility of the items.

As the area was examined in greater detail it was found that the parts were controlled in two separate ways. Some were controlled using a database, which had been installed by an external supplier. This gave the supplier access to look at the current stock levels and so re-supply as required. This meant that they were the main supplier of all the tool crib items. Any items that they didn’t stock as standard were still supplied through them and so were charged at a premium. It became clear that the company thought all items were supplied by this company but in reality only a small percentage were ever supplied to the tool crib, as the others were never used. When an item was used it was booked out of the database and so the stock count was kept accurate.

The items that weren’t held on the database were controlled using a card system. A card was held in the storage area and as an item was used the card quantity was adjusted. Written on the card were the minimum and maximum quantity levels. When the quantity level reached the minimum then the card was removed from the storage location and put on a desk at the front of the tool crib. These cards were then given to the buyer for them to order more stock. If an item was required that wasn’t in the tool crib then a purchase request would be produced. This item would then be ordered without any further questions asked. This could be anyone creating this supply request and so there was no control over it.

All the items were split into 40+ groups. These groups were determined by where they were used in the manufacturing process. The problem that became apparent with this approach was that it was found that the same item could exist several times as it would be categorised in several groups because it was used on different parts of the shop floor.

A location numbering system already existed in the tool crib but was not exhaustive and as the years had passed had become a little haphazard as further items were added to the inventory. It was decided to use the same method as what was used in the rest of the warehouse.

Before the data gathering exercise was started it had to be decided what information would be retrieved. A lot of the information on the database and card system was not very accurate and so it was decided to physically examine each item. This meant that manufacturing data could be collected, not just the supplier, and would help the purchasing department look for more competitive vendors in the future.
The information gathered was as follows:

- Lead-time
- R.O.Q. (Re-Order Quantity)
- Last used
- Control mechanism (card system / database / nothing)
- Location
- Our part number
- Manufacturer
- Manufacturer part number
- Vendor
- Vendor part number
- Price breaks
- Description
- Min / max levels
- Cycle time
- Category
- Price / unit

After this information was gathered it was found that over 3000 items had not been used for over one year and of these, over 1000 items had never actually ever been used. They were put into stock when the machinery was first installed on the shop floor and were items that the manufacturers had recommended that the company keep as spares. Even though these items had not been used for a long time it was still felt that it should be the department who may have to use them one day to be the ones to recommend disposal. To this end a list was sent to maintenance.

Adjustments to the descriptions of the items had to be done before it was migrated into the ERP system. This was because of the limitations in the description field. It was limited to 35 characters. The old systems weren’t limited in this way and so additional miscellaneous data had been added to the descriptions. It was thought at the time that this additional data could be used in a part of the new system called “characteristics”. The idea was that the item would be given a very basic description with all additional information in characteristics.

The next step was to decide on the basic description. The maintenance and stores departments did this together as there had always been a disagreement as to what each item was called. As this information was fed back it was used to populate the spreadsheet that had been created to gather all the tool crib data.

As part of the tool crib there is a KanBan area, which is sealed off from the rest of the tool crib. This has in excess of 200 containers. The maintenance and manufacturing departments have access to this area but not to the rest of the tool crib. Before the ERP system implementation the quantity level in this area had to be manually examined each day to make sure they didn’t run out. After first phase Go-Live this was handled the same as the other areas, with a minimum level. When this level is reached they are re-stocked by the rest of the stock in the tool crib. This is an improvement, as they all don’t need to be checked each day, only the ones that are highlighted by the ERP system.

3. PURCHASING

This department is responsible for the timely delivery of raw materials, project stock and tool crib items as demanded by the production plan. This department is split into two areas. These are procurement and project procurement. Project procurement is obviously concerned with the ordering of project specific items. More information on this area can be found in paper. [7]

The general procurement department is concerned with the ordering of raw material and consumables in the tool crib. The raw material part was quite easy to implement in first phase as there isn’t a great deal of variety to
the material and there are only a few suppliers. The biggest change here was that purchase requests and orders are now done electronically. When a part is delivered it can be matched to a purchase order on the system.

4. IMPROVEMENTS IN THE FIRST PHASE OF ERP IMPLEMENTATION

Many improvements have been made since the first phase Go-Live but it has been quite difficult to assign resources as the implementation of the second phase marches on. However some have been made.

In the tool crib it was found a lot of space was taken up by one small item. Most of the storage areas were of the same size and it didn’t depend on the size of the item. Where possible when the new location numbers were set up if the item was small then more than one location would exist on one shelf and so save a lot of space.

Using the new ERP system it became possible to look at the usage of the parts. The most frequently used items were then moved to locations that were nearer the issue area.

After the tool crib data gathering exercise was complete it became possible to create a list of items that were never used. This list was given to the maintenance department for them to examine. Even though some items had never been used in 10 years they still insisted in keeping some of them because they were classed as critical items because if ever they were required then production would stop and so cost a lot of money. Despite this a few hundred parts were disposed of as they were found to be obsolete. This would not have been possible without the data gathering exercise. More information regarding 5S can be found at [8]

It is now possible to have the ERP system create a purchase request when the minimum level for the item is reached. This minimum level is calculated on the quantity that is used on a daily basis, cycle time, and the lead-time to re-furnish the stock. If calculated correctly and efficiently this will be JIT.

Though this isn’t a true JIT system it is still reducing inventory. A true system is a pull system and reduces inventory, finished goods, work in progress and raw material. [9] More information on cycle time and lead-time can be found at [10].

It is possible, using the ERP system, to calculate safety stock, make a demand forecast and determine a re-order point for each item based on the demand history [11].

For the first phase tablet computers had been purchased to aid the use of the ERP system in the warehouse areas. Due to training issues and some teething problems with the wireless networks they were not used after the first phase implementation. Subsequently stores personnel had to go back to the office to interrogate the system. After approximately a month when people had got used to the new system a further push to train and demonstrate the system on the tablet computers was made. Initially this was met with stiff resistance from the stores personnel but as senior management consistently pushed this it was eventually successful. Receiving, issuing and stock take of inventory can now be done without entering the office and so a large time saving is made.

An extra function of the tablet computers is that they have a barcode reader built in. Each storage location and item label has a barcode on it. This saves a lot of time when inputting this information into the system as they can now be read using the reader. This was the main selling point to the stores personnel when training them to use the tablet computers as the tablets are smaller than standard and don’t have a full size keyboard.

A problem highlighted earlier in this paper regarding the way the project stock was ordered to a specific pipe to maintain traceability has now been rectified. Instead of ordering it for the exact pipe, it is now ordered to the project. The serial number of the pipe isn’t noted until it has actually been fitted to the pipe and so creates much more flexibility.

Now that procurement has a lot more visibility in the new system they have started an exercise to find and rate new vendors. They no longer want only one supplier for an item and so are now looking at several for each. It is now possible in the new system to create several suppliers for each item, each with differing prices and lead times.
5. IMPROVEMENT IN ERP IMPLEMENTATION KNOWLEDGE

There are some lessons that were learnt during the first phase. Some of these lessons are:

- **MAPPING**: It was decided for second phase to map the processes, rather than the departments, as it became obvious that processes affecting more than one department caused some of the problems in the first phase. The mapping for the second phase became much more thorough doing this.

- **COMMUNICATION**: It was found that it is very important to update the whole team and the departments affected by the ERP implementation, with the plans, issues and progress of the ERP project. In second phase this is now a priority as in first phase it was found that a lot of data was gathered, especially in the tool crib, that wasn’t used. It may get used in the future but time was very tight to make the Go-Live date.

  It was also found that to keep the organisation updated generally on the ERP project was beneficial as it helped to combat negativity. A monthly newsletter is now produced for the second phase.

- **TRAINING**: The timing and quantity of training should not be underestimated. When personnel are not using the system on a daily basis then it is easily forgotten. In the first phase training was done at an early stage but this didn’t seem to be beneficial as most of it was forgotten.

  In the second phase an IT literacy survey has already taken place to assess the present knowledge of IT systems and so help with the plan for training.

6. CONCLUSION

A lot of improvements have actually been driven by the ERP implementation and would probably not have occurred, or discovered, if it wasn’t for it. It certainly wouldn’t have happened so soon.

The second phase is now well under way and includes the implementation of the Planning, Commercial and Manufacturing departments as well as full implementation of Engineering and Finance. Engineering had an interim solution for first phase and so integration will be improved in second phase.

A lot of lessons were learnt in the first phase, some have been highlighted, and it has changed the way that the second phase is been implemented and it also helps that the ERP team have a good idea of what to expect and what problems can occur.

REFERENCES


Embedding ERP in an Engineering Projects Environment

Robert Lee*, Michael Boyce, Dr. Ahmed Abbas and Dr. Nasreddin Dhafr

School of Science and Technology, University of Teesside, Middlesbrough, TS1 3BA, United Kingdom.

ABSTRACT

Software solutions such as Enterprise Resource Planning (ERP) systems are contributing greatly to growth, efficiency and effectiveness of the business by providing end-to-end best practices. ERP system implementation in organisations requires understanding of strategic, tactical and operational levels of the business to embed best practice improvements. Hence, embedding industry best practices as an end-to-end solution is a common avenue within modern day manufacturing organisations to increase competitiveness. This paper focuses on the challenges involved in embedding an ERP system and associated processes within an engineering projects organisation. Changing market conditions, economic pressures and business improvement needs are explored to understand the strategic requirements for investing in large ERP implementation projects. However, consideration should also be made to the failure rates and impact on the host company in pursuit of this strategic change. This study puts forward observations during a phased implementation of an ERP system within such a bespoke manufacturing environment. Lessons learnt particularly in Projects and Engineering Department illustrate the challenges faced during the first phase in the ERP lifecycle encompassing the all important ‘Go Live’ date and post ‘Go Live’ activities.

Adaptation of the ERP system to the business requires using tailored tools and techniques such as process mapping, Business Process Re-engineering (BPR) and configuration management were employed to facilitate the improvement process. The tailored tools contributed to better visibility, acceleration of the implementation and improved traceability of parts, with more clear roles and responsibilities. Over and above this, the work also addressed challenges such as user acceptance, training, time constraints, resources, software limitation and communications with different functional areas. The activity undertaken specifically in the Projects and Engineering Department addressed issues with regard to standardisation, data integration, quality of data, reporting mechanisms and traceability. After overcoming the difficulties encountered in these areas, lessons learnt were created and stored as a knowledge base to enhance performance for the forthcoming phases of the implementation and minimise barriers on host companies while moving from a traditional company towards the ideal of a totally integrated enterprise.

1. INTRODUCTION AND BACKGROUND

Organisations are constantly looking for ways to gain a competitive advantage over their competitors. It is a core business necessity to remain competitive that is motivating and driving organisations to re-evaluate how they operate commercially. In recent years the world has seen the rapid improvement of information and communication technologies (ICTs) and the growing importance of Enterprise Resource Planning (ERP) systems as a critical resource.

This paper starts by reviewing historical events that led to the growth of Enterprise Systems, specifically focusing on the influence from business process re-engineering (BPR) in 90s and how this led to a surge of ERP becoming an enabler to initiate change. Furthermore the paper will comment on literature concerning the integration of ERP systems and implications on procedural and cultural issues across organisations. Using evidence from previous case studies we explore different views and perspectives, highlighting the important facts when implementing an ERP system. This is followed by a case study carried out at a small to medium sized company based in the UK.

* Corresponding author: Tel.:+44 (0)1912 959 621; E-mail: Robert.Lee@wellstream.com
The company specialises in bespoke manufacturing of equipment used in the oil and gas sector. The authors will follow the company as it embarked on an ERP implementation project from pre-implementation activities to Go-Live and beyond. This study concludes by addressing the challenges and lessons learnt from this case study as a result of integrating ERP in an Engineering Project Environment.

**BUSINESS PROCESS RE-ENGINEERING (BPR):** One technique that assisted organisations to transform how they worked in order to add value to the customer and dramatically cut costs was BPR. This section will briefly summarise the growth of BPR starting first with the business process perspective. A business process perspective simplifies complex tasks by splitting the entire procedure into smaller processes or steps. Process improvement and simplification is gained by examining each process or step to incrementally identify and eliminate bottlenecks and non-value events. [1]

In contrast to the improvements by increments, an alternative was offered suggesting it was;” better to throw a broken process away altogether and start from scratch than to improve it incrementally.” [2] This approach focused on radical re-design of processes which ushered in the new era for BPR.

Business process re-engineering (BPR) became infamous by the following statement. "Forget all you know...Don't automate, obliterate.” [3] This endorsed the view that any non-value adding processes would be “obliterated” instead of automating them using IT systems. BPR sole focus on value adding processes ultimately became its demise as it failed to take account of people factors. Evidence was seen of this in 1995 as “Pacific Bell announced that it was cutting 10,000 employees because of ‘reengineering’.”[4]

**ENTERPRISE SYSTEMS:** After the collapse of BPR, not surprisingly organisations wanted to consolidate their lost BPR activities. Considered by some to be the replacement, an Enterprise system was a commercial off the shelf package that was described as “the seamless integration of all the information flowing through a company”. [5]

Enterprise systems were originally designed for automating repetitive tasks. The emphasis of such systems was to establish a standard input, process and output to computerise the process in order to provide enhanced services and more functionality. By embedding generic business processes that were based on best practice and integrating them into all functional areas, enterprise systems looked like an ideal solution. However, “even though the promise of ERP is to offer a moldable configuration based on business requirements, current technology still inflicts a separate structure.” [6]

A report from a case study supported this claim and found the enterprise system forced inflexibility “with features that were at odds with its own requirements.” [7] Another case study noted that by standardising practices involved un-needed disruptions to many silos of activities.” [8] The main outcome of many implementations was the,” proliferation of failures in implementation and growing dissatisfaction.” [9]

Our research indicated, “the notion of ‘best practice’ can only ever mean the dominant perspective.” [8] In other words, process was no match for politics and the influential members of the board who exerted their power to change adopted practices. In this example, it is shown the importance for gaining the trust of people with power and status.

The Totally Integrated Enterprise (TIE) framework concurs with this view by stating, “Commitment and sign on is required from all functional units, else bottlenecks will occur.” [10] TIE’s emphasis is based on drawing together the ERP system, organisational functions and people with the aim of bringing deep rooted, and long term change through the process improvement activity.

**PROJECT ENVIRONMENT:** Due to the changes in customer requirements, many ERP vendors had to cater for demand from the growing project-based market. Within a project-based or engineer to order (ETO) environment organisations had the situation where products were designed, engineered and customised in pre-production to match each customer’s order. [11] As a consequence, maintaining a detailed list of all variations to the process became heavily taxing.

With the background set we have gained knowledge of the changes in the business environment that promoted the growth of enterprise systems and are presented with the following facts:

- Organisations have been reengineering their business process in order to achieve a competitive advantage since the early 90’s
- Enterprise systems are integrated solutions offering best practice. Although different scholars have reported Enterprise systems may constrain enterprising behaviour.
- Enterprise systems have moved to encompass ETO environments, and;
• TIE is a framework that brings together all aspects of the ERP systems and is used as a way to instill improvement.

2. CASE STUDY

Although implementations of Enterprise Systems have been documented within large organisations research is still in the initial stages for small to medium sized enterprises. This study was used to gain an understanding of an ERP implementation in a medium sized company attempting to embed ERP in an Engineering Project Environment. The company operates in a niche market where they design and manufacture bespoke products for the oil and gas sector. The host company appreciated that this was a huge undertaking which prompted them to seek professional consultancy and knowledge transfer from Teesside Manufacturing Centre (TMC) †. Before the ERP vendor selection stage The TMC used multiple tools and techniques including TIE concepts discussed earlier as the way of identifying improvements to make.

To understand an organisation in the context of this paper we will briefly introduce the organisational tiers, functional silos and the project environment. An organisation is a complex network of management tiers and information silos. An organisation can be categorised into three distinct levels. The strategic level, the tactical level and the operational level shown below (see, Figure 1).

![Organisational Tiers](image)

Figure 1: Organisational Tiers [12]

The strategic level includes the decision making activities with a long term view of the company’s future; these strategic activities can be further broken down into tactical activities. The tactical level which is comprised of medium term planning activities aims to achieve goals set about from the strategic level. The last level is the operational level which consists of day to day shop floor activities. The advantage using this approach is that it provides a structured way of thinking about information to be collected for management of an effective enterprise system. An organisation is also made up of a number of smaller self-governing functional silos. Organisations are often localised into different functional areas such as Engineering, Finance, Stores and Purchasing. (see, Figure 2).

† TMC is based in the School of Science and Technology at the University of Teesside. TMC specialises in areas including business process improvements, lean manufacturing, computer aided engineering and ERP implementations in small to medium sized enterprises.
This case study results from participant observation throughout a phased implementation of an ERP System and focuses on Engineering and Purchasing. A phased approach meant the implementation was completed in stages. Each phase of the implementation would develop a solution for one or more functional areas that would be integrated into the enterprise system before progressing onto the next phase. The breakdown of the phases was as follows:-

Phase 1. Purchasing, Stores, Finance and Engineering (partial).
Phase 2. Engineering, Manufacturing, Human resources, Quality, Sales, Document Management and Planning
Phase 3. Maintenance

All activities of the implementation of the enterprise System were governed by a Steering Group. The steering group consisted of the executives and senior managers from the company, consultants from the ERP vendor and TMC. The steering group was in charge of the strategic management of the project, and its first task was to form an ERP project team (see, Figure 3).

This consisted of a project manager and personnel from key functional areas including Stores, Engineering, Finance and finally two representatives from Purchasing. Consultants from the ERP vendor and the TMC were
booked to offer specialised advice when needed. Two high caliber graduates who were on the UK government supported Knowledge Transfer Partnership scheme (KTP Associates) were employed to provide additional support in the implementation. Having the assembled the project team, the following sections will discuss the process of embedding the Enterprise System.

**Embedding Enterprise Systems:** Enterprise Systems are implemented using three essential types of knowledge to embed them. The strategic management thinking is captured from staff at the organisation; best practice is captured from the consultants and best usage of the software from the ERP vendors. This trinity of knowledge needed to be balanced to achieve harmony in processes. Managing this balance can become complex and it was necessary that the project team aligned the business processes with their organisational tiers and functional silos. This meant that during the implementation period, the project team had to ensure that all function silos were not working in isolation. This made it possible to collaborate and standardise processes between function silos offering best practice across the supply chain.

**Enterprise Systems and Re-engineering Processes:** During the design stage stakeholders had to make another significant decision: should they try to re-engineer their processes before implementing ERP, or should they implement the ERP system first without undertaking thorough reengineering of their current processes? Option 1 was to commit time to reengineer processes before a major implementation, Option 2 was to integrate ERP whilst at the same time manage the business change process, or Option 3, which is to implement without changing the processes. Without streamlined processes the implementation would be encumbered with illogical and often non-repeatable processes. A decision was made to proceed with Option 2 whereby the integration and process change would be performed together.

In partnership with TMC, the company commenced mapping sessions to illustrate all staff duties at every organisational tier and functional silo. Process maps were created in collaboration with the personnel that had operational and management roles in the activities to capture the process fully. Using these current state process maps the team identified bottlenecks and any non-value adding activities. The team then embarked on a BPR exercise to agree to an ideal state map that facilitated the implementation process by enabling staff and ERP consultants to highlight possible improvement points and eliminate waste in the business process.

To commence putting theory into practice the ERP consultants configured the system ready for the implementation team to start testing. Each ERP team member would train their respective departments after the solution was finalised. It was found that the enterprise system overcame many issues based on standard jobs because it literally fits into an enterprise system designed for tasks such as bulk purchasing and receiving inventory en masse etc… Further evidence of this can be seen in [11] based on the same ERP implementation which shows benefits from a stores perspective.

**Lessons Learnt:** This section focuses on lessons learnt from the ERP system implementation and hopes to provide some valuable insight for companies preparing or in the midst of implementation proceedings. Although the embedding of ERP provided benefits, it was not echoed throughout all functional areas. One department that underwent major changes to totally integrate the organisation was the Engineering department.

Historically the Engineering department benefited from lots of flexibility. For example, in order to procure parts, Engineers would simply populate the relevant cell within a customised excel spreadsheet called the Bill of Materials (BOM). After entering minimal information such as the part description, the detail of the drawing and corresponding notes it was sent to the procurement department to buy.

With the introduction of the ERP system an inflexible process was enforced. For instance, in order to procure a part, engineers needed to acquire a unique part number, use standardised part descriptions and select valid tracking options which were traditionally managed by Quality. Moreover, the BOM spreadsheet which was now output from the Enterprise system created even more tasks for the engineers to perform as the Enterprise system needed to be instructed how parts related to each other to enable Procurement and Stores to buy and receive goods to the correct project.

The role that engineering played in the process had been radically altered and as a consequence produced areas where a greater effort was required to configure parts that would be accepted by the system rather than performing value adding engineering tasks. This raised the legitimate question about how much integration was enough? It was noted that the integration process caused discontent to other silos of activity.
To summarise the benefits of enterprise system provided:

- One single technology platform to share common data across the entire organisation.
- Real time sharing of information.
- Integration of function silos.
- Eliminate entry of duplicate information; and
- Indirectly optimised processes during BPR exercises.

Disadvantages are:

- Difficult to implement due to major changes to business practices.
- Inflexibility of integrated design; and
- Dependency on external consultants; this was an area of increased cost during implementation.

The ‘Go Live date came and passed without any major faults. However, overcoming the concerns noted pre Go Live was of top priority. In order to assess which improvements to take on first the company used a tool called the ease/benefit matrix (see, Figure 4). All concerns were noted and given a corresponding number. They were then rated by consensus from consultants and team members and mapped on the map below.

Figure 4: Ease / Benefit Matrix

Obviously solutions providing the most benefit and easy to implement took top priority while solutions that were hard to implement and offered lower benefits were given a lower priority. Members of the steering group and the project champion were actively involved in this stage to lend support where needed. Phase 2 of the implementation plan is currently underway whilst continuous improvements activities are implemented in parallel.

3. Conclusion

In a tough economic climate it is the authors’ view that Enterprise systems play an increasingly important part on the sustainability of businesses to perform at world class levels. During the work at the case study company there was an expectation of some business change but not at the scale initiated when embedding ERP. In order to
Implement the ERP system in the project-based company, the participants had to customise and configure the system to follow their own bespoke processes.

Customisation of an enterprise system within an ETO environment is much more time consuming and complex when compared to a mainstream implementation. Bespoke processes found only in the ETO/project sector created much frustration. As we saw examples of this in the lessons learnt section where the system forced engineers to prepare overly detailed BOMs thus making a fairly simple task in creating an engineering part turn out to be a complex chore. That is the limitation of ERP systems at the moment. The whole aim of ERP systems is to find the optimum solution – which is fine for volume manufacturing but not so good for project-based environments. While this and other technical issues will be addressed, the author does not envision this being available without modifications.

Nonetheless, there was evidence of the company benefiting from ERP systems when receiving bulk parts in. As standard parts could be created en masse and only needed to be configured once in the system. The main reason for this was that enterprise systems work most easily when configured to fit with volume or batch jobs. Although ERP systems have become increasingly powerful there is still no one-size-fits-all solution to cover all the differences between companies that are volume-based and project-based. Unfortunately, this means chances are higher of encountering problems with customisation of processes or software limitations in an ETO environment.

However, technologies will continue to develop and ERP systems will become more pervasive in the project environment. Future research will conclude with the next phase of the implementation and should provide quantifiable results to analyse the success of the ERP implementation so far. The recommendation by this paper is for organisations to work together as a whole and connect its people, processes and technologies by doing so will aid in developing an optimum solution for all.

REFERENCES

Design of a Generic Methodology Framework for Process Diagnostics Management

Jiri Tupa*, Chike Oduoza

Department of Technologies and Measurement
University of West Bohemia in Pilsen
Univerzitni 8, CZ 306 14 Pilsen, Czech Republic

School of Engineering & the Built Environment
University of Wolverhampton
Schifnal Road, Priorslee, Telford, TF2 9NT, UK

ABSTRACT

Process management tools and techniques applied to manufacturing process environments are essential for enterprise continuous improvement. Recent research have studied methodologies and made recommendations for application of the process management to manufacturing, information or administrative processes. The ITIL library is an example of very sophisticated solution for information technology service management (IT). It documents industry best practice guidance focusing on providing high quality services, while placing special emphasis on customer relationship in the use of IT services. However, there is no evidence in the literature of a suitable process management tool for diagnostic purposes in the system / process life cycle. There is, therefore, a need to design a generic methodology framework for process diagnostics management. The aim of this paper, therefore, is to develop a methodology framework for diagnostics applications in manufacturing processes management. The methodology used for system design were developed based on detailed diagnostics processes analyses and personal interview of researchers and industry practitioners. The diagnostics framework contains steps leading to process management implementation and techniques for planning, execution and evaluation of the diagnostics processes. The framework integrated the ITIL library principles. The ITIL is structured to address problem and incident management for service support, and describes procedures which can classify data as incidents and problems. The framework is designed for quick and easy decision-making.

1. INTRODUCTION

Manufacturing and service enterprises use diagnostics systems for preventive and maintenance management. The main purpose is to test products during manufacture processes and also to monitor the production system. The diagnostics system is an important tool for management set correction and improvement of processes. Diagnostics management covers planning, realization, execution and improvement of diagnostics processes. The objective of diagnostics process management is to achieve as well as guarantee quality assurance.

Recent research and application have methodologies and recommendations for application of the process management in manufacturing, information, maintenance and administrative processes. The ITIL library is an example of sophisticated solution for information technology service management (IT) and documents industry best practice guidance focusing on providing high quality services, while placing special emphasis on customer relationship in the use of IT services. However, there is no evidence in the literature of a suitable process management tool for diagnostic purposes in the product / process life cycle. There is, therefore, a need to design a generic methodology framework for process diagnostics management.

* Corresponding author: Tel.: (+420) 377634531; Fax: (+420) 377634502; E-mail: tupa@ket.zcu.cz
2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

ISO 9000 for quality management and ISO/IEC 17025 which is the standard for management and technical procedures refer to the requirements for quality management system implementation and certification. These standards can be used for management system establishment in diagnostics management. ISO/IEC 17025 is focused on the operation and effectiveness of the quality management system within the laboratory, and its technical requirements to address the competence of staff, methodology and test/calibration equipment. Laboratories use ISO/IEC 17025 to implement a quality system aimed at improving their ability to consistently produce valid results. It is also the basis for award of accreditation. Since the standard is about competence, accreditation is simply a formal recognition of a demonstration of that competence. ITIL (Information Technology Infrastructure Library) is the documentation of best practice for Information Technology Service Management and consists of a series of books giving guidance on the provision of quality IT services, and on the accommodation and environmental facilities needed to support IT [1] - [4].

ITIL has been developed in recognition of organizations' growing dependency on IT and embodies best practices for Information Technology Service Management. The problem and incident management is part of service support management. The goal of Incident Management is to restore normal service operation as quickly as possible and minimise the adverse impact of incidents on business operations, thus ensuring that the best possible level of service quality and availability is maintained. “Normal service operation” is defined here as service operation within Service Level Agreement (SLA) limits. Incident Management must be as quick as possible. An incident is defined any event which is not part of the standard operation of a service and which causes, or may cause an interruption to, or a reduction in, the quality of the service.

The main goals of Problem management are: to minimize the adverse impact of incidents and problems on the business; to prevent recurrence of incidents related to these errors. The problem management process requires the accurate and comprehensive recording of incidents in order to identify effectively and efficiently the cause of the incident trends. The main advantage of ITIL application is the complexity and depth of solution and ITIL books cover following management areas: security management, software asset management, financial planning, strategic management etc. The solution which is included in the ITIL standard is based on “best practice” recommendation.

The disadvantages are presented. Sometimes ITIL is used by manufacturing managers who do not understand this standard, because ITIL describes processes and management process for informatics system but not for manufacturing systems. Manufacturing managers understand only the quality management standards of ISO 9000. Also definition of scope application and translation ITIL requirements for manufacturing can be a problematic.

Although ISO and ITIL standards are support business process management application the relationship between ITIL and BPM has described suggests support ITIL with process-oriented tools such as workflow management systems by Branner [4]. It discusses the need for workflow management support of service management processes to achieve service level compliance, and presents criteria for determining which IT Service Management processes can and should be supported by workflow management systems. The IT Service Management processes defined by ITIL are evaluated and divided into four basic process classes according to their suitability for workflow management, thereby laying a foundation to future top-down approaches for comprehensive ITIL tool support.

2.1. BUSINESS PROCESS MANAGEMENT

One way to achieve process performance management system establishment is by successful implementation of business processes. Management of business processes is an approach used in many industrial applications [1]- [12].

A high number of citations for example a Google.com search on „Business Process Management“ returns more than fourteen thousand pages where this phrase appears. The science on Business Process management BPM issue is the subject of research focused on methodological or technological solution of BPM problems [1]. The main problems of applications BPM are described by Wasana Bandara in article [5], where fourteen global experts were interviewed. The outcome form this study describes the lack of governance, standards, methodology, tool support for process visualisation etc in the implementation of BPM.

3. METHODOLOGY DESIGN

Designed methodology concept helps to implement the diagnostics process management system based on ITIL application. The concept has been developed into two steps:

1. Analysis of diagnostics processes and current needs of responsible managers, staffs, researchers and technicians in diagnostics area. The designed questionnaires were used in this step. We have obtained process attributes of diagnostics processes.

2. Study, selection and modification of suitable methods and tools for business process management and performance measurement.

3.2. METHODOLOGY FRAMEWORK

Methodology framework is presented by Table 1. As we can see the table presents key steps and activities in defined order. The methodology concept is based on business process and process performance management theory and the implementation of ITIL. It means application of BPM and PPM in first step. This step leads to:

- process analysis and key process indicators setup
- process description, modeling and optimization,
- system of evaluation and process execution

The second step ITIL application helps to mark the outputs of diagnostic process as incidents or problems e.g. ITIL describes reference processes for support of decision making. The outputs from this process can be data or information about diagnosed object (diagnosis). Implementation of information infrastructure namely knowledge databases for incidents and problems help to reduce the time for decision making and information analysis.

<table>
<thead>
<tr>
<th>Step</th>
<th>Activities</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy definition and key process and performance indicator set-up</td>
<td>Analysis of current situation Definition of mission, vision and strategic goals.</td>
<td>Key processes indicators are defined</td>
</tr>
<tr>
<td>Process analysis and mapping</td>
<td>Process mapping Process classification Process attributes description</td>
<td>Processes are analyzed.</td>
</tr>
<tr>
<td>Process modeling</td>
<td>Implementation of process modeling methodology Process models making</td>
<td>Processes are described.</td>
</tr>
<tr>
<td>Process optimization</td>
<td>Definition of main optimization criteria Process optimization</td>
<td>Processes are optimized.</td>
</tr>
<tr>
<td>Problem and incident management implementation</td>
<td>Establishment of processes and rules for incident and problem management, implementation of knowledge databases.</td>
<td>The problem and incident management is applied according to ITIL.</td>
</tr>
<tr>
<td>Process measurement and system for key process indicators execution and evaluation</td>
<td>Design of process measurement and execution system Determination of periodicity of process measurement Implementation of tools for process measurement, execution and evaluation Design of system for correction proposal and improvement</td>
<td>System for process evaluation and execution of processes is implemented. Improvement system is defined.</td>
</tr>
</tbody>
</table>

The information infrastructure can also be built on Service Oriented Architecture (SOA) which provides methods for systems development and integration where systems group functionality around business processes and package these as interoperable services. SOA also describes IT infrastructure which allows different applications to exchange
data with one another as they participate in business processes. Service-orientation aims at a loose coupling of services with operating systems, programming languages and other technologies which underlie applications [10].

3.3. CASE STUDY

The case study presents the designed methodology framework application for diagnostics of soldering in electronics production process. The soldering process is an important part in electronics assembly. The aim of diagnostics process is to ensure quality control of soldering material and components. This kind of process is an example in off-line diagnostics which is used before soldering process starts for process setup or in case of breaking production due to unacceptable failures in process. It is releasing to determine what happened and why in a short time. Due to this fact the diagnostics process has to be standardized and described. The practical application of designed methodology presents how this process is described and managed.

The methodology implementation is now described as follows.

3.3.1. STRATEGY DEFINITION AND KEY-INDICATORS SET-UP

The aim of this step is to define key process indicators based on performance metrics definition. In this case the analysis of current situation has been done from discussion with specialists from the diagnostics area. The specialists identified these current problems:

- diagnostics processes aren’t described,
- processes aren’t monitoring and executed,
- information and knowledge aren’t shared,
- system for diagnostics process management doesn’t exist.

These problems are analyzed by applied process management system according to the designed methodology. Firstly the strategic goals and key process indicators were formulated. The main strategic goals were formulated in four perspectives of Balanced Scorecard. The key process indicators have been developed for process execution and process performance measurement. The key performance indicator types can be divided into 3 groups:

- time-related key performance indicators (e.g. throughput times, processing times, frequencies),
- cost-related key performance indicators (e.g. process costs/rates on the basis of the performance standard) and
- quality-related key performance indicators (e.g. number of processors, error rates, deadline reliability).

<table>
<thead>
<tr>
<th>KPI</th>
<th>Type</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval of order processing</td>
<td>Process</td>
<td>Time</td>
<td>Process duration from an order acceptance till issue of an invoice.</td>
</tr>
<tr>
<td>Interval of acceptance of an order</td>
<td>Process</td>
<td>Time</td>
<td>Sub process duration: “Receipt of order”.</td>
</tr>
<tr>
<td>Interval of testing soldering</td>
<td>Process</td>
<td>Time</td>
<td>Sub process duration: “Soldering testing”.</td>
</tr>
<tr>
<td>Activity process time</td>
<td>Function</td>
<td>Time</td>
<td>Difference between the end of current activity (process) and the end of previous activity.</td>
</tr>
<tr>
<td>Processing time of an activity</td>
<td>Function</td>
<td>Time</td>
<td>Difference between the end of activity and the start of a particular activity.</td>
</tr>
<tr>
<td>Idle time of activity</td>
<td>Function</td>
<td>Time</td>
<td>Difference between the start of an activity and the end of previous activity.</td>
</tr>
</tbody>
</table>
3.3.2. PROCESS ANALYSIS AND MAPPING

The diagnostics process of soldering was strongly analyzed in this step. The aim was identification and description of all processes attributes and activities during this process. Process mapping means description of all processes attributes and relationships between them. The result of process analysis is demonstrated in Table 3.

Table 3: Process attributes

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Name</td>
<td>Soldering testing</td>
</tr>
<tr>
<td>Goal</td>
<td>Soldering testing of selected parameters of electronics devices or printed circuit board.</td>
</tr>
<tr>
<td>Owner</td>
<td>Mr. Novak</td>
</tr>
<tr>
<td>Customer</td>
<td>External</td>
</tr>
<tr>
<td>Input</td>
<td>Order and sample</td>
</tr>
<tr>
<td>Output</td>
<td>Diagnostics Report</td>
</tr>
<tr>
<td>Key process indicator</td>
<td>Process time</td>
</tr>
<tr>
<td>Sub-processes</td>
<td>Sample preparation, Soldering test</td>
</tr>
<tr>
<td>Output information</td>
<td>Standards ISO or IPC requirement, Test setting, Information about material and testing samples</td>
</tr>
<tr>
<td>Equipment</td>
<td>Testing soldering machine</td>
</tr>
<tr>
<td>Costs</td>
<td>Material costs, Personal costs, Energy costs</td>
</tr>
<tr>
<td>Energy</td>
<td>Electricity supply power network 230 AC</td>
</tr>
</tbody>
</table>

3.3.3. PROCESS MODELING

We decided used ARIS methodology was used for process modeling. This methodology is based on oriented object modeling languages. ARIS allows describing reality from another point of view. The main model - model of added value as presented in Figure 1 is one example. It is basic view on process structure and relationship for top managers. The three kinds of processes can be seen in this model. They are control (planning, marketing), main and support (quality management, accounting, service) processes. The model can be decomposed as a detailed process entitled eEPC (Extended Event-driven Process Chain) process model in ARIS. Figure 2 presents example of a test soldering model. The key attributes of this process are recorded in this model.

Figure 1: Generic Diagnostics Process Added Value Model
In this step the ITIL Problem and Incident Management was applied. The aim of this application is definition key process for problem and incident solving. Figure 3 shows the process model of knowledge database formation. In this model we can see the main input to database. It is Incident and Problem report (outputs of previous and incident management processes). Figure 4 presents the main event and process in incident management. The start is rise of incident on diagnostic process. The incident has to be delivered to Internal Customer Contacts Diagnostics support. Detected incident has to be recorded to Incident Report. The data from report are compared in knowledge database and the incident is marked like a new incident. In case of known incident the data in database can be updated. Next processes described in Figure 4 are started on based of the result of detection process.

Figure 2: Model of soldering diagnostics

3.3.4. PROBLEM AND INCIDENT MANAGEMENT IMPLEMENTATION

In this step the ITIL Problem and Incident Management was applied. The aim of this application is definition key process for problem and incident solving. Figure 3 shows the process model of knowledge database formation. In this model we can see the main input to database. It is Incident and Problem report (outputs of previous and incident management processes). Figure 4 presents the main event and process in incident management. The start is rise of incident on diagnostic process. The incident has to be delivered to Internal Customer Contacts Diagnostics support. Detected incident has to be recorded to Incident Report. The data from report are compared in knowledge database and the incident is marked like a new incident. In case of known incident the data in database can be updated. Next processes described in Figure 4 are started on based of the result of detection process.

Figure 3: Process model of knowledge database formation
The report contains the description of problem or incident identification. The reports contain predefined and compulsory form item. The database provides in information platform and these records are stored in the knowledge database supported by IT platform as part of enterprise information system. The output of the knowledge database can be a model of causes and events. Service team can use this result for support for decision-making about incident and problem solving.

3.3.5. Process Measurement and System for Key Process Indicators Execution and Evaluation

The last step is implementation of system from process performance management. In this case the system is based on process control application. Processes control involve measuring efficiency of the business processes implemented in IT systems and putting control systems in place that will monitor compliance with range of legal regulatory requirements. It means building of information architecture, system and implementing software tool. This solution corresponds with the new trend in process management.

Companies that are able to align their processes to the requirements of their environment and surroundings will not only gain a competitive advantage, but they will also be able to manipulate this alignment better and faster than their competitors. The prerequisites for this solution are the supply of decision-relevant information and an ability to transform this information quickly and effectively into sustained measures for targeted alignment of business processes. We decided to use the ARIS Process Performance Manager (ARIS PPM) software tool for process control in this case, because this tool can be implemented to any information system and structure. Owing to the fact that we are using tools ARIS for process modelling and optimization of processes, we decided to use the software ARIS PPM.

First experience of ARIS PPM implementation has shown benefits of this solution in the manufacturing area for technological process control, in this case for diagnostic process management. The SW tool ARIS PPM helps to make analysis and processes monitoring, particularly:

- real process course,
process time,
- comparing of the real and planned key indicators value,
- type and kind of order.

The results of process mining and analysis have been used for process models correction and KPI planning. The responsible management obtains quick management and performance view.

4. CONCLUSION

Diagnostics processes, information and knowledge represent important values for a company these days. These process attributes should be controlled if the business process management is implemented. This paper presents design of methodology for diagnostics process management and shows one way to manage information and knowledge in ITIL. ITIL (Information Technology Infrastructure Library) has been described in previous and the example of the application was given in the case study.

The case study focused on diagnostics in the manufacturing area. The results of the study present a model of the soldering process and short application of ITIL. The example of description implementation of the ITIL for diagnostic of the soldering process was also shown. The results can be used in information systems implementation and provide a sophisticated base for responsible management and decision-making. The first experience of implementation indicates that designed solution helps to reduce of process time and control of process quality.

A complex analysis of designed methodology and ITIL application for diagnostics process management would be one possible direction for further work.

ACKNOWLEDGEMENTS

This research is supported by the research plan MSM4977751310 “Diagnostic of interactive processes in electrical engineering”.

REFERENCES

Global Manufacturing: Creating the Balance between Local and Global Markets

Parisa Gilani 1 and Yasamin Razeghi 2

1 Independent Marketing Consultant
Bournemouth, Dorset, BH22 9RL, England

2 PhD Research Student
Shahid Beheshti University
Tehran, Iran

Abstract

For several years the domestic markets of manufacturing organisations have started to reach maturity [1] and companies have sought to expand their international operations in order to grow. This has meant there has been an increasing emphasis on the debate on whether companies should remain global or localise their marketing mix, and to what extent each element should be adapted or standardised. International marketing can be defined as ‘the performance of business activities designed to plan, price, promote and direct the flow of a company’s goods and services to consumers or users in more than one nation for a profit’[2]. In this paper the authors argue against the statement ‘manufacturing organisations are either mindlessly global or hopelessly local’ and demonstrate that manufacturing organisations can successfully combine a global and local approach if they carefully choose the elements that they adapt or standardise.

1. Definitions of Adaptation and Standardisation

The process of “globalization,” refers to the big-picture process that draws products, services, and markets around the world closer together. It is a process that involves a complex array of actors and institutions, including firms, governments, NGOs, and consumers. This process is typically analyzed at the macroeconomic level, where the country is the unit of analysis.

While academic and social debates about ‘globalization’ ‘global village’ etc continues, corporations are expected to create profitable businesses and generate returns for investors by entering trans-border markets, compete against international rivals, risk investments, find and develop opportunities in a turbulent world which is in continuous competitive transformation. Their power of leadership and foresight needs to manage extended international enterprises, enter diverse international markets, and manage risks and uncertainties that range from global supply chains to financial risks to geopolitical risks.

Production is increasingly a global activity. In this ongoing internationalization of production, multinational corporations (MNCs) have played a critical role. While most multinationals still originate from advanced, developed nations, there is a growing number of MNCs from countries that do not belong to this select group.

There is a range of international activities that takes place within the framework of the multinational corporation and these activities can be analyzed from a number of different points of view.

Globalizing therefore refers to the process by which a given firm begins a journey of becoming global in its vision and corporate objectives in order to achieve competitiveness. Few, if any, firms have completed this journey ‘truly global’; many, however, are globalizing. Globalizing is a process that unfolds not at the level of the country but at the level of the firm, and consists of the actions firms have to take as they become more engaged in that process.

An approach that is prevalent in a large part of the MNC literature takes the internalization decision as given and focuses primarily, and in more detail, on the division of labor between parent and affiliate within
the MN C. MNCs th at relocate activitie s internationally o rganize p roduction both in sco pe (the set of products) and in depth (stages of production) between the parent and the affiliate in such a way as to take advantage of international differences in factor abundance and factor prices [3], [4] and [5]. Interestingly enough, the theory of the multinational corporation as initially developed is primarily conceived from the perspective of advanced, developed countries [6], [7]. These theories primarily view the MNCs as a vehicle to relocate labor-intensive activities to affiliates in low-wage countries. They rationalize on one-way MNC activity in which MNCs from the capital-abundant North set up affiliates in the labor-abundant South [8], [9].

To understand the process of globalizing, we certainly can not ignore the broader process of globalization. Our global political and economic systems create the context for our business enterprises. As corporations with affiliates abroad, MN Cs al most by de finition hav e a vertical component. A M NC’s affiliates abroad need firm-specific (intangible) assets such as headquarter services, R&D, patents etc. to operate and these will be traded within the firm. Since this latter vertical component is part of any MNC, Markusen and Maskus [10] do not make it the distinguishing characteristic between, what have been called, vertical and horizontal integration models of the MNCs. Essential for the vertical models in their view are differences in factor intensities of the components of the production process and relocating these to take advantage of factor price differences.

The MNCs from these countries have affiliates in both more and less advanced countries. On the one hand, the liberalizations in China, Malaysia, Thailand or Vietnam account for the boom of affiliates from MNCs with parents in these middle-income countries.

There is no all-encompassing theory of the multinational corporation. There are a variety of models that focus on different aspects of multinational activity that all have some empirical relevance. MNCs take advantage of factor prices and have them break up production accord ingly. On the other hand, other researchers argue that breaking up production to take advantage of factor cost differences need not be an essential motivation for MNCs [11], [12], [13]. Access to rich markets is key and to save transportatio n costs, MNCs may set up shops in other developed count ries instead of hav ing to sup ply these markets through exports, even without factor price differences. Finally, there is the view that MNCs emerge in an effort to jump tariffs or other existing costs associated with trading goods [14].

The degree to which manufacturing organisations are and should be either globally or locally orientated and therefore the adaptation and standardisation debate is by no means a new subject of discussion, with increased attention from the late 1960s. Buzzell [15] was one of the first academics to address the concerns facing companies wishing to operate on a national scale. He argued that standardisation has the potential to be a useful approach, but there are barriers to this. In this way multinational standardisation implies the same product, the same price, the same distribution systems with the same promotional techniques and message in every country that the firm wishes to operate in [15]. This would relate to the idea of ‘global’ above. In contrast a ‘local’ marketing strategy would be one in which no elements of the marketing mix are the same in any one national context. The interesting and relevant point to make about this piece of research is that at it s disavantages both it s and disavantages to s standardisation and therefore adopting a global approach. However it fails to examine exactly which elements of a marketing mix should be adopted. We can now examine in further detail the standardisation side of the debate.

2. THE STANDARDISATION ARGUMENT

Levitt [16] argued that there is no need to adapt any elements of the marketing mix to suit local tastes due to the ‘globalisation of f markets’. Therefore he would argue that manufacturing organisations can be global in their marketing approach without being ‘hopeless’ in fact that a global approach is in deed a successful one. Levitt [16] can be identified as the principal advocate of the standardisation viewpoint. He argued that the world is becoming more homogeneous with technology being the driving force behind it. A firm that was in accordance to this viewpoint would standardise all elements of its marketing mix.

There are numerous advantages for a firm choosing to adopt a standardised approach. One such advantage is that standardisation allows a brand to build up a strong global presence that is easily identifiable [17], and thus consistency with customers. A second advantage is the ability to take advantage of economies of scale [18], which in turn leads to significant cost savings, particularly through product design. One early example of this is Henry Ford’s Model T where he stated that you could have ‘any colour
so long as it’s black’ [19]. Thirdly a standardised approach can improve planning and control and finally standardisation allows a company to exploit good ideas that are formed in one context and extend these to other contexts [18]. However evidence within the business context heavily suggests that there are very few manufacturing organisations if any that can be identified as taking a completely standardised approach.

There are also disadvantages to adopting a largely standardised strategy. Firstly a standardised strategy is primarily product driven and fails to take into account the needs of consumers [20]. This could lead to disastrous results if the needs are very different from the benefits that a product supplies. For example Chee and Harris [18] state that consumer tastes and preferences are very specific and vary considerably. Furthermore if the products are not adapted and therefore remain at the same cost to produce, the price of the products could be so that they do not match what is affordable in a given country or more expensive than local competitors.

Furthermore there are many cases in the business context of companies who have tried to introduce their brand to international markets, without adapting. One such example is Hallmark. Hallmark are known in the U.K for their greeting cards with sentimental messages within, so that the British customer does not have to write their own message. However when Hallmark tried to introduce their brand to the French market they did not adapt their product apart from translating the messages into French. The French prefer to write their own messages in greeting cards and therefore the Hallmark brand failed in France [21].

3. THE ADAPTATION ARGUMENT

The polar opposite of standardisation is adaptation. Advocates of this viewpoint highlight that there are important differences between countries in terms of culture, stages of economic and marketing development, political and legal systems and customer values and lifestyles [2]. This would imply that adaptation is the appropriate approach. Kacker [22] distinguishes between compulsory (legal) modifications and voluntary modification. He claims that marketers can increase the probability of success by being aware of and responding to different cultural realities. We can thus infer that adaptation has the benefits of responding to customer demand in each cultural setting. Douglas and Wind [20] state matter of factly that arguing that a strategy of universal standardisation should be applied ‘appears naive and oversimplified.

However in recent times there has been an increasing tendency to adapt some elements of the marketing mix and standardise others. Simon-Miller [23] states that the only global brands are McDonalds, Kodak and Coca Cola. However the authors would suggest that even Coca Cola adapts to local tastes and changes product names accordingly. For example in France what is termed Diet Coke in the U.K has been renamed Coke light in France to meet legal requirements and country specific expectations. Diet Coke would imply to this culture that instead of having reduced sugar content the drink would aid weight loss. Additionally McDonalds at one point was seen as a true global brand. However it adapts to local market conditions and customer market needs. For example in France McDonalds expands its product offering to include Beer and yoghurts. It is evident that different companies adapt or standardise to different extents. There are many authors who have attempted to identify to what extent elements of the marketing mix should be standardised. The EPRG Paradigm examines this process through examining management orientations.

4. THE EPRG PARADIGM

One such approach is examining management orientations through the EPRG paradigm. According to Keegan [24] the way that a company responds to global market opportunities is dependant upon the way the management sees and understands the nature of the world. This ‘worldview’ is explained through the EPRG Paradigm. The first orientation that management can adopt is the ethnocentric orientation, which is also termed the market extension approach. This is defined as assuming that the home country or domestic market is ‘superior’ compared to the rest of the world. In this way companies who see the world in this manner assume that products and practices that are successful in the home country will succeed anywhere in the world without adaptation. This is type of strategy may lead to economies of scale and lower costs, however as a result sometimes opportunities outside of the home country are ignored and the company fails to understand the local culture.
By contrast the polycentric orientation can be defined as the ‘belief or assumption that each country in which a company operates is unique’ [24]. It can also be termed the multidomestic approach. Companies adopting this approach identify the differences in each national market and treat each one as unique. In this manner large scale adaptation occurs. This approach has the advantage of being able to understand the local culture and adapt accordingly. However it is costly and if adaptation is carried out on a large scale the company can lose brand identity and a consistent brand image across each of its markets.

A Regiocentric approach is one in which product uniformity is sought within a cluster of markets [24]. The cluster can be organised in a variety of ways including geographic proximity, language or membership of a regional union. A company adopting this type of approach benefits from economies of production and marketing. However if the chosen market cluster is too large it could slow down the ‘diffusion’ of the product. One example is if the company decides that its market cluster is going to be the European Union. The European Union is expanding constantly and now encompasses a large variety of cultures within it.

Finally a geocentric approach is one in which the company views the world as a single market [18]. Within this approach there is a high level of centralisation and coordination of activities which in turn leads to lower research and development costs, rationalised product lines, economies of scale and worldwide distribution of the product. However this strategy is largely dependent on careful and continuous market research. This however is not the same as an ethnocentric approach; it is instead a combination of an ethnocentric approach and a polycentric approach in that it is a global strategy that is highly responsive to local demands. Companies that adopt this type of approach are often termed transnational companies, with Unilever being an example of success.

The different orientations are summarised below. Arrows numbered X represent advantages and arrows numbered Y represent disadvantages.

<table>
<thead>
<tr>
<th>Ethnocentric or Market Extension</th>
<th>Polycentric or Multidomestic</th>
<th>Regiocentric</th>
<th>Geocentric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host country is ‘superior’</td>
<td>Each host country is unique</td>
<td>Sees similarities and differences in a global region</td>
<td>Worldview</td>
</tr>
<tr>
<td>Missed opportunities</td>
<td>Understand Culture</td>
<td>Slow ‘diffusion’</td>
<td>Lower Costs</td>
</tr>
<tr>
<td>Economies of scale</td>
<td>Costly</td>
<td>Dependent on market research</td>
<td></td>
</tr>
<tr>
<td>Economies of production and marketing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Economies of scale

Missed opportunities

Understand Culture

Dependent on market research

Slow ‘diffusion’

Lower Costs

Figure 1: Adapted from Howard Perlmutter [25]. The addition of the arrows are authors own contribution.

Following one of these paradigm approaches determines to a certain extent the level that a company chooses to adapt or standardise its marketing mix to each international market. It demonstrates to some level that manufacturing organisations are not necessarily ‘mindlessly global’ or ‘hopelessly local’ and in
fact that some companies adopt a strategy that is somewhere in the middle of this spectrum. Nevertheless an ethnocentric approach might be portrayed as ‘mindlessly global.’ The framework is a clear and useful tool for examining an enterprise’s view of adaptation and standardisation, but it is not prescriptive and once these orientations are evident within a particular company they maybe hard to change. In fact Perlmutter [25] stated that even if a firm wishes to have a geocentric orientation there are both obstacles and drivers to developing this orientation such as the distrust of big international manufacturing organisations by political leaders in host countries. This would be a significant obstacle to a geocentric orientation. The framework looks at the organisation and its management as a whole rather than looking at each element of the marketing strategy. However the EPRG paradigm is not the only approach in existence. Many authors have sought to provide an integrated standardisation framework.

5. FRAMEWORK FOR STANDARDISATION

What is perhaps interesting and advantageous compared to the previous approach is that Jain [26] distinguishes between marketing program and marketing process. Program refers to the different elements of the marketing mix and process refers to ‘tools that aid in program development and implementation’. According to Jain, the degree of program standardisation or adaptation depends on five individual factors. The first of these factors is the target market.

5.1. TARGET MARKET

According to Jain, the decision on whether to standardise or not is situation-specific, which is evidently a different way of examining the issue than the EPRG paradigm. Target market encompasses both the geographic area and economic factors. For example, economic factors influence the level of development of a country and it can be argued that companies with their home base in a more developed country moving to a country with a similar level of development are more likely to standardise than those from a similar domestic market moving to a country with lower economic development. Indeed, when Chandra, Griffith and Ryans [27] attempted to adapt Jain’s framework to transferring advertising from the USA to India, they found that similarity between the target markets of the USA and India would be associated with more program standardisation. Chandra, Griffith and Ryans state that as a partial result of the similarity between the American and Indian markets manufacturing organisations have been successful in standardising their advertising programs.

5.2. MARKET POSITION

The second important factor to consider according to Jain [26] is that of market position, which in turn encompasses market development, market conditions, and competitive factors. If a product’s foreign market is at a different stage of development than that of its domestic market, it is likely to need to adapt to match this gap. The three conditions that affect the degree of standardisation or adaptation that is required are cultural differences, economic differences, and differences in customer perceptions. There are many examples of companies who have failed to take account of cultural differences. One such example is Kellogg’s whose cereal brands include Corn Flakes and Rice Krispies. In the 1990s, after increased pressure for competitors and stagnant U.K. and USA markets, the company decided to expand its operations to India. The decision was based on the low pressure competitive environment and the size of the population. In 1994, after international barriers to trade in India had been taken down, Kellogg’s decided to enter the market. However, due to cultural differences, the concept of eating breakfast cereal in India was new and novel. Kellogg’s not only had to market their own brands and product lines but also the concept of eating cereal. This proved to be unsuccessful as the Indian consumers treated Kellogg’s products as a one-off purchase. However, from an other perspective, it was an extent to meet cultural differences can damage the brand identity [21]. According to Holt, Quelch and Taylor [28] one of the three factors that affected consumers’ purchasing decisions with regards to global brands is ‘global myth’. According to this piece of research, consumers see global brands as symbols of ‘cultural ideals’ and see them as a way of creating ‘an imagined global identity that they share with like-minded people.’

5.3. NATURE OF THE PRODUCT
The third element which is considered to be important in determining the level of standardisation is the nature of the product, which encompasses the type of product and product positioning. Indeed Boddewyn, Soehl and Picard [29] found that the degree of standardisation that is feasible depends on whether the goods are con sumer non-durables, con sumer dura bles o r i ndustrial goo ds. St andardisation for con sumer non-durable goods such as foods standardisation would be most difficult as a result of differences in national tastes. This is perhaps interesting as there are a variety of examples of food brands that have attempted to use standardisation across national borders such as Mars for example. In the case of consumer durables such as washi ng machines or ca meras standardisation can be at an advanced level, with only slight adaptations made to meet leg al require ments. However it is pre dicated th at the most a dvanced standardisation will occur with industrial products. Additionally it is thought that if a brand or product is positioned in the same way in all overseas markets standardisation is feasible to a large extent.

5.4. ENVIRONMENT

The fourth important element is the en vironment. This is sig nificant as it consid ers fac tors from a n external perspective. This element includes physical, legal, political and marketing infrastructure factors. This piece of research supports the seminal article by Buzzell [15]. According to Buzzell different countries have different legal require ments with product design, advertising, competitive positioning, pricing and employment. We can see evi dence for example of the different restraints on advertising throughout the world. In Kuwait only 32 minutes of advertising are allowed per day, advertising to children is banned in Sweden, alcohol advertising is banned in certain countries as is tobacco advertising [2]. So me of the most significant political constraints are tariff and other trade barriers, local content or export require ments and the existence of cartels. According to Jain [26] the marketing infrastructure ‘consists of the institutions and functions nece ssary to create , develop and service demand.’ Buzzell [15] states that the differences and availability of certain types of media and or distribution channels can affect the degree of standardisation. Perhaps the most relevant concern with media today is that of the internet, which has become a powerful source of advertising within the USA and the UK. In the USA there are 501 internet users per 1000 people where as in China for example there are only 11 internet users per 1000 people [2]. So we can see here that at the legal level compulsory adaptations are needed, but the core marketing can be kept the same and voluntary adaptation might be necessary with relation to media reach limitations.

5.5. ORGANISATION FACTORS

Organisation fac tors en compass corporate o rientation, headquarters-subsidiary r elationships and del egation of auth ority. Cor porate orientation addresses the issues o f the EPRG par adigm that were analysed earlier. As we can see this corporate or management orientation forms only a small part of Jain’s standardisation framework and in that way this framework is more comprehensive. The importance of the headquarters-subsidiary relationship and the delegation of authority coincide.

6. SUMMARY

Product standardisation was very popular in the 1980s, but since then the focus has shifted to adapting at least some elements of the marketing mix. It appears that standardisation arguments are built on a series of assumptions about consumers that are not necessarily always accurate. There are various frameworks that exist to determine the level of adaptation that is necessary by a company wishing to market its products in international markets. From this research and examples of companies that have been both successful and failed dramatically when entering into new markets it appears that the most appropriate response is to maintain a core product and add peripheral modifications. The advantage of this is that the company can keep a consistent global brand image for example like Coca Cola or McDonalds that has a good reputation and a certain personal ity and co notations at tached to it, however it can adapt according to legal requirements and customer preferences like tastes in every culture that it operates in. The strategy that is adopted should depend on the product and the target market. Evidently no such strategy is guaranteed success.
REFERENCES

Potential of Benchmarking

Munir Ahmad*

School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

This paper describes how to measure and benchmark manufacturing performance as a first step on the journey to performance and competitiveness improvement. Based on the research and experience we have simplified the benchmarking process and have developed a methodology which can be easily applied by an organisation in order to identify the potential benefits and hidden plant. It has been quite common to determine in companies during the first benchmarking exercise that; the hidden plant may well be in access of 30% of output, stocks may be reduced by 50%, the fixed cost can be reduced by anything up to 20%, potential variable cost reduction by a further 10% and customer service is poor. These potential benefits identified through benchmarking provide interesting opportunities, given that they may often be achieved with minimal to zero capital expenditure, and are delivered through the people of the organisation itself. The measurement framework of benchmarking a process is covered based upon three elements; measuring the output of the process which is the performance, measuring the input practices that are applied to deliver these outputs and measuring the culture and softer issues that ensure that the practices are applied and outputs are delivered. This paper provides a methodology which has been applied to various manufacturing companies but does not offer a prescribed template for benchmarking and scoring performance. While it provides the framework, measures and industry world class targets, the actual detailed framework and scoring must be determined by the individual company. A set of data and case studies are also presented to help the organisations understand how to interpret the mass of information generated in a benchmarking study and focus on priorities to start the improvement process.

INTRODUCTION

Manufacturers are required to provide their products to their customers at the expected time, with reasonable prices, and meeting the required quality [1]. Benchmarking is a performance improvement tool and usually used to improve process effectiveness, product quality and service delivery. It enables an organisation to compare existing performance to others, and identify elements that can be adopted in their business context. The result is often a business case for making changes in order to make improvements. Benchmarking is widely used, and there are many different approaches; ranging from models of quality measurement to basic comparison undertaken on a pragmatic basis. It is often necessary to tailor an approach that provides a sustainable method of benchmarking in view of the organisational goals and business needs. There are four types of benchmarking [2]:

- **Internal benchmarking**: this type of benchmarking applies more to large organisations that have various business units. It is about comparing internal operations from one sister company to another.

- **Competitive benchmarking**: this specifically compares competitor to competitor, using the product or function of interest. This goes beyond the traditional product or service engineering, but it is important to learn and understand about competitors’ methods, their processes, innovation, strategies, markets, etc.

- **Functional benchmarking**: this is the comparison of similar functions within the same broad industry, using wide industry leaders as partners.

- **Generic benchmarking**: this involves establishing the comparison of business functions or processes that are the same, regardless of type of industry. Generic benchmarking is a long-term challenge and reflects a total change in the culture. It is usually found in learning organisations.

The benchmarking often uses a number of key performance indicators (KPIs) for comparisons of performance. The KPI is a number, or value, which can be compared against an internal or external target ‘performance benchmark’, to give an indication of performance. That value can relate to data collected or calculated from any process or activity [3], [4]. The KPIs are often classified as; ‘strategic-in-nature’ KPIs, or ‘operational-in-nature’ KPIs. Operational KPIs are used by managers to plan and control programs at operational level, while strategic KPIs provide guidance to both managers and policymakers who have to make decisions from a more global perspective. Generally, KPIs should be selected according to how they relate to the

*Contact details: Tel.: (1642) 342443; Fax: (1642) 342401; E-mail: m.m.ahmad@tees.ac.uk
strategic aims of the organisation and how they can support the corporate-level performance indicators. This means that KPIs, whether they are operational or strategic in nature, must tie in to strategy.

The first task for the persons undertaking the benchmarking is to use their experience and mature judgment of improvement processes, plus their knowledge of the plant, to determine what is practically achievable in the future and what the priorities are for that particular plant. For example, while the hidden plant may represent the largest financial opportunity, the plant operates in a low cost restricted market so the focus will be to reduce the costs.

**THE PROCESS PLANT’S BENCHMARKING METHODOLOGY**

The methodology of benchmarking the process plants as shown in Figure 1 was developed [5] to assist in the performance improvement of the process plants. The manufacturing performance assessment introduced by this methodology is presented in Table 2.

---

**Figure 1: Methodology of benchmarking the process plants**

In this methodology the authors have presented a framework on how to measure and benchmark the performance of the process manufacturing plants. They have defined the following measures for use in benchmarking the process plants:

1. **Customer services**
   - On Time In Full (OTIF)
   - Customer complaints.
   - Due date reliability.
   - Adherence to production plan
   - Stock turn.

2. **Reliable assets**
   - Product rate
   - Quality rate
   - Availability

3. **Operational excellence**
   - Statistical process control
Potential of Benchmarking

- Manufacturing velocity

4. Motivated people
- Absenteeism
- Training days
- Staff turn over

5. Safety, health and environment
- Annual reportable injury accidents
- Environmental performance

The sequence of these KPIs, however, gives a complete set of strategic performance measurements, related to different strategic dimensions. The appropriate combination of measures is determined by the nature of the actual industry. A bulk chemical industry, for example, may focus more on the adherence to sales and operational planning (S&OP), whereas a consumer product industry is encouraged by its final customers to pay particular attention to issues such as OTIF and customer complaints. In general, these KPIs support the corporate performance indicators at the corporate level of the organisation, such as ‘return on investment’, ‘return on assets’, ‘market share’, ‘customer loyalty’, ‘stakeholders satisfaction’, and so on. It was also recommended [5] to use fewer focused measures to make the process more effective. As a guideline it was advised that the total number of measures should be less than 15.

The benchmarking targets are decided and made available once the type of benchmarking has been agreed upon [5]. It was stated [6] that many companies have a problem when performing benchmarking activities. This problem is caused by a difficulty in retrieving quantitative data relevant to world-class or competitors’ performances. It was also highlighted [6] that there is little work being done to address this problem or, at least, to propose a dynamic technique that can help manufacturing organisations in this area. When data for benchmarking are not available, the organisation is advised to establish targets the best way it can [6]. Organisations have always set performance targets and used measurement systems to track the achievement of targets [7], [8], [9], [10]. There are four methods that can be used to set or generate performance targets [7], [8]. They are as follows:

- **Historical targets:** this method is often used, owing to its simplicity in implementation [11]. The best historical performance of an activity would be defined, then used as a performance target.

- **External targets:** by using this method, the world-class performance of best-in-class operations is collected, to use for benchmarking purposes. Table 1 shows some superior performances of process industries’ world-class plants that are often used as performance targets. Importing external benchmarks from competitors’ performances, however, is often an essential requirement for competitiveness in any business.

Table 1: World-Class Performance of Process Plants (Source: [5])

<table>
<thead>
<tr>
<th>No.</th>
<th>Key Performance Indicator</th>
<th>World Class Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adherence to Production Plan</td>
<td>&gt; 99 %</td>
</tr>
<tr>
<td>2</td>
<td>Overall Equipment effectiveness (OEE)</td>
<td>&gt; 95 % for continuous plants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 85 % for batch plants</td>
</tr>
<tr>
<td>3</td>
<td>Process Capability (CpK)</td>
<td>&gt; 2</td>
</tr>
<tr>
<td>4</td>
<td>On Time In Full (OTIF)</td>
<td>&gt; 99.9 %</td>
</tr>
<tr>
<td>5</td>
<td>Stock Turn</td>
<td>&gt; 25</td>
</tr>
<tr>
<td>6</td>
<td>Value Added per Manufacturing Employee</td>
<td>£400K</td>
</tr>
<tr>
<td>7</td>
<td>Training Days per Employee</td>
<td>12 days</td>
</tr>
<tr>
<td>8</td>
<td>Absenteeism</td>
<td>&lt; 1 %</td>
</tr>
</tbody>
</table>

- **Internal targets:** these targets are based on comparisons of the performance of internal processes. In this method, the performance of the best performing process in the company will be used as a benchmark. This method is myopic, however, in the sense that it looks only within the organisation and may limit the company’s performance relative to its external environment [11].

- **Theoretical targets:** in this method, the targets are determined by the use of theoretical calculations, or by simulation analysis. Setting performance targets on a theoretical basis is exceptionally useful, as it defines the performance capability and limitations of the process. A considerable disadvantage of this method relates to the difficulty and time needed to define it. The theoretical targets of the process would enable effective comparisons of process performance to benchmarks from external sources; they would help in
assessing the capabilities of operations and whether or not they are able to perform similar to best-in-class operations.

Some calculations on how to define performance gaps in the process manufacturing plants can also be used [5]. Some of these calculations are as follows:

Variable cost gap = \((\text{actual variable cost} - \text{world class variable cost}) \times \text{output}\)

Fixed cost gap = \((\text{actual fixed cost} - \text{WC fixed cost})\)

Stock savings gap = \(\left(\frac{1}{\text{Actual stock turn}} - \frac{1}{\text{World class stock turn}}\right) \times \text{Turnover}\)

Hidden Plant = Output \(\times\) \(\left\{\frac{\text{World Class OEE}}{\text{Actual OEE}} - 1\right\}\)

\text{OEE: is the overall equipment effectiveness measure.}

A benchmarking report [12] revealed that the competitiveness of the UK process industries is average when compared with world class levels and that there is a ‘hidden plant’ of 28% of the UK’s current effective capacity. That is, if the UK’s average OEE performance was raised from its current level to the world class levels then the UK process industries’ effective capacity would increase by 28% for minimal capital expenditure. Releasing this ‘hidden plant’ would support additional sales of nearly £10b per annum which increases value to shareholders.

Additionally, a lean manufacturing study [13] highlighted that for most production operations only 5% of activities add value; and 35% are necessary non-value activities; while 60% of activities add no value at all. This study highlighted the fact that the elimination of waste represents a huge potential in terms of manufacturing improvements and the key is to identify both waste and value; develop knowledge management base; and realise that sustainable improvement requires the buy in of the people operating the processes and managing the business, and therefore a culture of continuous improvement.

The relationship between application of best manufacturing practices in manufacturing plants and the ability to achieve superior performance (see, Fig. 2) is usually studied and findings are used to drive improvements.

Fig. 2: Practice - performance relationship
OPERATIONAL PERFORMANCE MEASURES

Operational performance measures are in some cases selected for benchmarking, with respect to their relevance to the operational objectives and the strategic KPIs. Some of the commonly used operational measures are as follows:

- **Manufacturing cycle time**: the total time that the process takes, from start to finish, producing the required production. This time normally comprises of the time required for cleaning, loading, processing, and emptying the equipment.

- **Value of stock in warehouse**: all the money the system invests in purchasing things the system intends to sell.

- **Cost reduction**: KPIs under this category measure the achieved cost reduction through an improvement program.

- **Amount of wasted materials**: material losses from processing, quality control inspections, handling and transfer of materials.

- **Quality measures**: these measures comprise of ‘right-first-time’ performance, defects per unit of production, and detailed pareto diagrams, etc.

- **Equipment uptime**: uptime is defined as the time a process operates at its standard maximum proven rate, while making a defect-free product. The ratio of uptime in the total available time is often calculated as follows:

\[
\text{Uptime} = \frac{\text{Valuable operating time}}{\text{Valuable operating time} + \text{Total losses}}
\]

- **Equipment downtime**: downtime is defined as the period when the equipment is not available to fulfill its intended role. Downtime can consist of planned activities, such as scheduled maintenance, repair and overhaul activities and unplanned activity as a result of a failure, or accidental damage.

- **Resource utilisation**: resource utilisation is defined as the proportion of time that a resource is used.

- **Final product yield**: it’s the percentage of transformed raw materials into salable product.

- **Production capacity**: production capacity is defined as the maximum rate of output that a production unit is able to produce under a given set of operating conditions.

The basis for calculating operational performance measures, however, differs from one sector to another, from plant to plant, or even from one operation to another operation within the same plant. The manufacturing cycle time is an example, as, in some cases, it might be calculated by simple, straightforward equations, or in other cases, may require complicated calculations.

6. **BENCHMARKING IN A CHEMICAL PLANT**

This selected example is a multi-product, multi-purpose batch plant. It operates under a company that is a leader in engineering, production and after sales service in the oil sector in Africa. The demand forecasting activities in this plant are based on judgment and depend on experience. The output of the forecasting process defines the expected sales volume and this output is used to support planning decisions. The management of the company has been concerned for some time about the lack of performance measurement and lack of benchmarks that facilitate the planning decisions. The manufacturing assessment [5] is used to assess performance of the plant. Section 4 of this assessment is shown in Table 2.

The analysis of results revealed that the plant requires an improvement strategy which can take the performance of the plant to the world class levels. The overall performance of the plant scored 24 points (see, Table 3). And according to this performance the plant is in division 3 ‘Below Average’, therefore, according to the performance index attached to the assessment, the plant has scope for improvement. A detailed analysis of the operational and maintenance excellence is required to assess the potential improvements. Through the manufacturing assessment, the plant is at risk with the existing performance of OTIF (on time in full) delivery performance, that is, customers can use new suppliers once they find them. Any supplier has OTIF between 71 and 99.99 % is better than this plant’s performance. The plant also needs to adhere to its production plan by elimination of process variability that causes processes to take longer time, and also reduce downtimes caused by setups, and breakdowns. The plant has low OEE resulting from low product rate which is again resulting from process variability. The company has very poor performance in regard to stock turn; therefore, this measure requires maximisation through reduction of all inventories and increase of sales. The absenteeism and
training are people issues and both have direct impact in productivity and quality of production. The plant has low performance in the measures related to these two issues.

It was also identified that tools such as Kanban, Kaizen, just in time (JIT), statistical process control (SPC), business process reengineering (BPR), total productive maintenance (TPM), single minute exchange of dies (SMED), poke yoke and 5S are some of the most appropriate tools to improve manufacturing performance in this plant. In general, a sustainable performance improvement should be aligned to a firm's ability to continually achieve quality, reduce costs, and deliver products or services to customers in the expected time. However, performance problems can arise from different causes; therefore, there is no unique solution to solve all problems.

### Table 2: Performance Assessment

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Actual Performance</th>
<th>World-Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing Added Value/Employee</td>
<td>£36K</td>
<td>£400K</td>
</tr>
<tr>
<td>3.1 Manufacturing Added Value per Manufacturing Employee for the plant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 OTIF Delivery performance %</td>
<td>70 %</td>
<td>99.9 %</td>
</tr>
<tr>
<td>3.3 Adherence to production plan %</td>
<td>60 %</td>
<td>&gt; 99.9 %</td>
</tr>
<tr>
<td>3.4 Customer complaints % of orders delivered</td>
<td>15 %</td>
<td>&gt; 0.01</td>
</tr>
<tr>
<td>Reliability and Consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5 Product rate %</td>
<td>65 %</td>
<td>90 %</td>
</tr>
<tr>
<td>3.6 Quality rate %</td>
<td>95 %</td>
<td>99 %</td>
</tr>
<tr>
<td>3.7 Do you routinely measure lost time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheduled downtime - % of capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unscheduled downtime - % of capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability is 100% less the sum of scheduled and unscheduled downtime</td>
<td>90 %</td>
<td>95 %</td>
</tr>
<tr>
<td>Calculated OEE</td>
<td>56 %</td>
<td>85 %</td>
</tr>
<tr>
<td>Control Flexibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 Maintenance cost as % of replacement asset value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.9 Process Capability (CpK)</td>
<td>&gt; 2</td>
<td></td>
</tr>
<tr>
<td>3.10 Change time of routine tasks affecting production, including grade changes (in minutes)</td>
<td>60 minutes</td>
<td></td>
</tr>
<tr>
<td>Value of material and source stocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of work in progress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of engineering spares &amp; stores items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of finished goods stocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total value of stocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock turn</td>
<td>0.00005</td>
<td>&gt; 25</td>
</tr>
<tr>
<td>People</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.12 Absenteeism %</td>
<td>4 %</td>
<td>&lt; 1 %</td>
</tr>
<tr>
<td>Safety: All injury frequency rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average training days per employee</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Total Score</td>
<td>24</td>
<td>&gt; 80</td>
</tr>
</tbody>
</table>

After the analysis of performance, it was concluded that the plant requires some improvements to start at the operational level. The operational effectiveness model (see, Fig. 3) is chosen as a tool for improvement of the plant’s operational performance. The aims were, however, to achieve reduction in the total cycle time of some selected processes, optimise batch sizes, and increase yield of the final product.

The utilisation of manufacturing resources in this plant as in many other plants is a crucial issue. In an ideal manufacturing operation, the processes would operate at full utilisation of all resources. For example; there would be full utilisation of manpower, equipment time, and capacity, there would be zero waste from the handling and processing of raw materials combined with no scrap on the process utilities. However, this situation is rare in real operations. The difference between the ideal and the real situation is due to losses [7], [8].

### Table 3: Performance Index
The plant manufacturing has considerable scope for improvement. The plant requires a detailed operational excellence assessment, with a high likelihood of significant benefits through manufacturing, maintenance, and operational changes.

Division 1: Poor
The plant manufacturing performance has scope for improvement. A detailed analysis of the operational and maintenance excellence is required to assess the potential improvements.

Division 2: Below Average
The plant appears to have manufacturing and maintenance practices which, while average, could be improved. The assessment has highlighted some areas of the plant with further potential.

Division 3: Average
The plant has good manufacturing and maintenance practices and it appears to be exploiting learning opportunities.

Division 4: Good
The site is aware of the scope for further maintenance and manufacturing improvement. The plant is approaching a winning standard and demonstrating leading manufacturing and maintenance practices.

The operational effectiveness model is used to reduce the losses in the plant to the minimum levels. The manufacturing objectives of the plant were defined by informal interviews with some key individuals in the plant. It was concluded from the study that, the minimum production cycle time, maximum yield of final product, and minimum total production costs are the main three objectives of the plant. This was a good input for design of the operational effectiveness model.

There are several benefits of cycle time reduction and some of them are: minimum waste in time, cost reduction, increased flexibility, less time for risk probabilities on hazardous processes, and efficient employees.

The raw materials have the largest portion in the manufacturing costs. When the raw materials are used inefficiently, a waste will be created which will lower sales revenue and create an environmental problem [14].

The economic batch quantity is a best practice to achieve minimum cost of manufacturing. The capacity decisions are vital to the manufacturing plants as they are affecting number of dimensions in the business such as delivery date, product prices, and warehouse requirements. Costs associated with production setups, resource utilisation, and inventory holding are all related to the batch quantity and contributing to total manufacturing costs. The link between stocks and financial performance of organisations is strong [16]. The organizational objective ‘maximizing return on assets’ is a good example of this link. An advantage of manufacturing at economic batch quantity is the tuning of the output of manufacturing to either meet only the market demand, or to produce for demand and for inventory. Therefore, when the manufacturing process optimised for cycle time
to be the least possible time, and yield to be at its maximum potential level; then the economic batch quantity will ensure that the plant will not overproduce to meet the objective of those two factors.

8. CONCLUSION

The methodology used in this paper suggested that there will be a significant opportunity to improve the performance of the majority of the existing assets. Through benchmarking, gaps are identified and an outline of a procedure to quantify existing gaps was provided. All the tools provided by the methodology of benchmarking within the process plants focus on quantifying a financial gap; this sets out the priorities and justifies the case for continuous improvement. After identifying the gaps, guidelines were given for companies on how to introduce process improvements which are realistic and achievable. The paper also has investigated the techniques that used to define a performance targets to assist manufacturing plants in attaining their goals. No one target setting method is enough in itself as the competition may have deployed better equipment, technologies or innovations to generate high performance results. Therefore, it is advisable to use the different target-setting methods in conjunction, as none of these methods are adequate, individually, for determining performance targets. The results of application at the selected case study plant revealed that there are significant losses on time and money. The improvement in the time of the process will lead to an increase in the value of the product, thus creating financial savings.

REFERENCES

Moving towards Creating an Organic Lean Manufacturing Organisation

Professor Sadri Gilani\textsuperscript{1} and Yasamin Razeghi\textsuperscript{2}

\textsuperscript{1}Managing Director  
Rahyaft Consulting  
Tehran, Iran

\textsuperscript{2}PhD Research Student  
Shahid Beheshti University  
Tehran, Iran

\textbf{ABSTRACT}

In recent years there has been widespread discussion concerning the shift away from traditional production paradigm and clear advancement of lean manufacturing as a fundamentally new approach. This approach is dominated by product quality, differentiation, and price. The paper argues that the move beyond high performance production is necessary in order to create long term organization competitive advantage. Organizations are entangled in environments where producers, suppliers' customers, and external agencies interact with each other. This interaction stimulates and creates innovation and change in all the participants. Thus an organic environment is required to accommodate both the lean manufacturing and the internal organization sense making. The use of sense making concept as the main tool for organizational learning is pivotal. The paper illustrates how the above have been developed and implemented as an integrated approach to creating an organic lean manufacturing organization in two large manufacturing companies in a developing Country. The approach combines the use of Emergence Strategy approach, Lean manufacturing, sense making, capability enabling, change creation, and problem exploration and solving approaches.

1. INTRODUCTION: PAST AND PRESENT- LEAN-FIT AND NOT LEAN-THIN

One may credit Henri Fayol as a main pillar of modern management and theory with its core assumptions of a high performance organization in the early part of the 20\textsuperscript{th} century. He saw performance achievement in the context of an organization infrastructure that was centralized, functionally specialized and hierarchical, highly defined, in which everything had its specific place. Management was perceived as a champion for planning, organizing, forecasting, coordinating and controlling. Many of the later management researchers followed the same philosophy [1], [2] by proposing new tools for the same framework. Generally our current cost/waste reduction tools or frameworks still focus on improving the implementation of Fayole's view. In reality majority of management authors have provided an extended model of this view, which is still much in use to the present day.

2. ORGANISATIONAL FORMS FOR THE 21\textsuperscript{ST} CENTURY

In the last two decades more innovative views of organization structures started to emerge which valued the integration and synergy of human capability, creativity, and technological advancements. Traditional views that people should be told what is best started to break down. A de-structuring of the over-structured forms emerged in a number of industries [3]. In the face of increasingly competitive environments, the traditional mechanistic focus appears weak and breakable [4].

As organizations entered the concept of global village, they realized that global environments value response, flexibility, choice variations and not, pre-fixed and predefined products. As their customer base changed to a more educated and informed entity the organizations realized that customers can define their needs and expectations. As a reaction to this fundamental change a new paradigm for organisational forms and behaviors started its evolutionary cycle [3]. Organizations aimed that they have moved from efficiency to effectiveness. The lack of definition for effectiveness, and over generalization of this concept,
initiated a huge weakness for the future development of performance management. The wishful thinking for finding a solution to all problems suddenly became a mirage of reality with the introduction of 'Lean production', downsizing, and core capabilities. Organizations which previously had no clear reasons for getting rid of employees or closing down unprofitable divisions, suddenly found a plausible justification. Instead of developing new horizons for organizational and management thinking, they were again pushed down the Foyle 'le' mind-frame. As the global customer base grew and economic wealth increased, organizations faced large demand for their products. Organisations moved their production sites to low-cost countries. The implementation of this move and limitation of educated workforce in those regions meant the further justification of 'new tools and old framework'.

In a parallel cycle some organizations looked to different approaches and were not blind followers of fashionable cues. They moved on a change in design principles that amounted to a major shift which included success factors, innovation, creativity, integration, flexibility, and cross functional synergy [5].

The current global financial crisis seems to be a blessing in that it forces us to rethink our naive prides and over-claimed assumptions. Organizations were all moving so fast to solve all our problems using 'Lean and mean' organizations with guaranteed competitiveness and success. Lean was everything. Suddenly we are faced with the questions such as: why has VW group increased sales while Toyota (lean production initiator) did not. Why did GM, an advocate of LEAN face immense problems. There are no clear evidences that show our Lean organizations did better that the not-so-Lean organizations in the current crisis.

In the post-industrial/information age organizations seem to have followed the same principals of the Taylorism values using new techniques and methodologies. All the techniques and approaches such as business process reengineering, supply chain engineering etc... Push organizations to do better and better, the same old thing.

It seems that they are being brainwashed that the customers require the 'old things' but better and cheaper. The traditional Taylorism view of the company as a centrally-driven entity that creates wealth by getting better and better at doing the same thing is a till persist [6] in our traditional view of a 'lean' production. At the same time new concepts such as lean production, learning, and lean manufacturing are requiring us to think creatively within the context of achieving highly specified targets with minimum wastage and bare requirements. Flexibility and space for change as given in not really focused and assumed customer and producer behavior. The focus has changed from creative human capability building to maximizing the human resource productivity. The inregnation required between strategic management and engineering, and manufacturing paradigms is becoming superficial [7]. Continuous deep innovation is based on the ability to synergistically balance the above domains [a system that specializes in exploitation will discover itself becoming better and better at an increasingly obsolescent technology. A system that specializes in exploration will never realize the advantages of its discoveries] [8].

Competition in the intensive industries is no longer complex; rather, it is in the non-traditional, creativity-intensive industries that we face the real challenges. The future requires value creation by the conversion of 'possibilities' to 'customer desires'. Industries are on the beginning of a 'business des ign' and 'product creation' revolution in our organizations which enriches itself organically, and accepts that seeing future as an investment, must include some additional activities which in the short term is non-value adding activity. In the academic world, this approach was well practised such as industrial placement, consultancy, and research. Sadly enough even the academic world is moving towards the failed approach of the industrial world.

If businesses are to ensure long term competitive survival, they need to enrich Operational effectiveness with 'creativity competence effectiveness' and refine Strategic flexibility, with 'strategic change creation'. In this way we can see that the future successful companies need to focus on excellent exploration capabilities rather than exploitation mechanisms.

The criticality of organisation's structure, size, strategy, technology, environment, culture, and the ir complex and intrinsic relationships have been highlighted a number of researchers. Burns and Stalker [9] propose a fit between structure and the rate of change in its environment must occur if an organisation is to achieve maximum performance. Mintzberg [10] has extensively and passionately researched the importance for organisational structure. Miller [11] has exemplified the important links of strategy and structure. Handy [12], [13] suggests implication of cultural factors in relation to organisational design and structure and the need for new organisational forms. Pascale, Millem and Gioja [14] demonstrate the critical role of design approach and creativity as the life line of the organizations. Mabey, Sala man and...
Moving towards Creating an Organic Lean Manufacturing Organisation

Storey [3] identify the critical relationship between organisational structure and design, and human resource management.

During the 1950s, the Toyota Motor Company, inspired by the waste minimization concepts developed by Henry Ford in the early 1900s introduced advanced manufacturing methods that aimed to optimize the resources it takes for a single product to flow through the entire production process. Toyota engineered an organizational structure focused on the systematic identification and elimination of waste from the production process. James Womack, Daniel Jones, and Daniel Roos introduced the term “lean production” in their 1990 book *The Machine that Changed the World* to describe the manufacturing approach established by the Toyota Production System. In this context any activity that did not directly lead the creation of product was considered as waste.

Lean production typically represents a paradigm shift to reducing time intensity, material intensity, and capital intensity of production for product-aligned “one-piece flow” pull production.

In the concept of traditional LEAN how can corporate loyalty be Value added activity and how can it be related directly to main production activities. The traditional lean approach at most will allow organizations to relate this to their current well defined products (understood customer expectations). This seems to develop a culture in which organizations accept that rather than creating future desires we are addressing current customer expectations. This is a dangerous path to getting better at the old things.

3. AN ENRICHED VIEW OF LEAN FIT (ORGANIC LEAN) APPROACH

If one assumed that organisation performance is related to their ability to manage various kinds of social problems, and to deal with the motivational and emotional problems arising due to negative feedback and failures [15], [16] then we may look for a different integration of Lean Concept. This was indeed the authors approach in the creation a Lean-fit manufacturing capability in a number of companies in a developing world country.

In this approach two main classifications for creating a Lean-Fit organization were identified. A person enriched in both should be resilient and able to handle the commercially and emotionally threatening consequences of failure and success.

1. The first group is called Lean – factors including: lean operational tools and concepts. Tools and concepts in this category included high-performance work systems, high-performance work practices, total quality management, and target management.

2. The second group consists of eleven Organisational Strength factors which represented our focus from work organization and human resource management. These included: Sense-making capability, trust, Corporate loyalty forces, Integrated communication culture, Creativity and creative thinking, Shared experiential learning, Bottom-up Strategic capability and belief, Internal ly created change initiation and management, Shared achievement focus. It is important to note that since products are always changing the organization must replace product loyalty by corporate and customer loyalty.
4. ROLE OF EMOTIONAL INTELLIGENCE, SENSE-MAKING, COMMUNICATION, AND WORKGROUPS

Employees work as a way of developing social and economic status, and enjoy emotional attachments as a way of creating motivational energy, and value their attachment as a sense-making process. Their loyalty is a result of this process which may have no real rational logic or evidence behind it. It does however create surplus energy within the individuals in the company.

Sense-making is a process of continued redrafting of an emerging issue, which is enriched by further data through the process of exploration. The idea that sense-making is driven by plausibility rather than accuracy is in contrast with the assumption that the accuracy of managers’ perceptions determines the effectiveness of outcomes [17], [18], and [19].

Researches regarding the accuracy of manager’s perceptions have questioned the reliability of the manager’s perception [20]. The induction and deduction processes of turning abundant data into actionable knowledge seem to be the key [21], since sense-making plays a crucial role in this knowledge formation cycle.

Organizational sense-making is the creation of an environment which allows the employees to develop an agenda for change through the process of internalized group loyalty. The synergy created through the group sense-making will itself create confidence throughout the organization.

If one can connect the abstract with the concrete through experimentation then we can create an organization with can deal with crisis as a challenge with immediate actions, local context, and concrete
Moving towards Creating an Organic Lean Manufacturing Organisation

cues [23], and not a threat. Sense-making framework can again be a fundamental tool for the lean-fit organization.

As such one can argue that lean production in reality should be seen as the 'product workshop' culture and not a whole organization body. In that context lean must engage the organisational strength factors if competitive sustainability is to be achieved. If we were to assume this, perhaps we can then state with more confidence:

- Lean produces an operational environment that is highly conducive to waste minimization and focused production, which must be enriched by organisational strength factors in order to be self-creative and innovative. Lean methods focus on continually improving the resource productivity and production efficiency, which frequently translates into less material, less capital, less energy, and less waste per unit of production. OS factors will enable it to look beyond current requirement and explore a new pathway. Any activities which is supporting the creation exploration can no longer be assumed to non-value adding. We must therefore create three positions for all activities: direct value adding, indirect value adding, and non-value adding.

- Lean can be leveraged to produce more internal improvement, filling key “blind spots” that can arise during lean implementation. Their integration with OS factors will allow us to see external opportunities within internal improvement which allows us to create new markets, new products, and new human resource capabilities. Our people create tomorrow ahead of customer requirements.

5. SUMMARY

Lean manufacturing has helped organization competitiveness. We must however accept that its main focus is productive flexibility, process standardization, and systemization. It may be misleading to expend the concept of lean production as lean management or even lean organization. For organizations to survive the volatile global environments they must build in contingencies in their capability, structure, knowledge, and flexibility. It may be possible then to accommodate future business diversifications using the current workforce. The lean concept has been wrongly used as an excuse for outsourcing many capabilities which we need internally in time of change and economic uncertainty. It has also been an excuse for many senior managers for reducing workforce which do not directly create short term value. Investment in human resources has also been turned into training for short term compliance. Production culture has traditionally been slow to accept customer focused approaches. In the market place customer segmentations are not clearly separated any more. As customer find other means of creating buying power, companies must make sure their employees possess the same flexibility and innovation power. The current world economic crisis has shown the rigidity of our mechanistic views. In the two large manufacturing companies in the middle-east, the framework of lean manufacturing was enriched by the integration of work organization factors such as shared knowledge, cognition, and learning. This allowed and encouraged innovation beyond process refinement. Using the concepts of high-involvement management, and high-performance practices a further motivational framework was created. By in-depth integration of lean manufacturing, work organization, and human resource management an organic learning was achieved. The organizations were able to create an environment where high-involvement and commitment from management and employees are emphasized and required. In both cases not able results were achieved. Using the same workforce production targets rose by over 50%, the wastage declined by 67%, and employees suggested renovation and innovation of their processes and working systems. Delegation of decision making was seen as a crucial element of team-working, and a framework for sharing the benefits was developed. In this way the development of an organic lean manufacturing organization where the culture of commitment replaced compliance took place.

REFERENCES


The strategic importance of ERP screening process: Evidence from the forging industry

Athanasios Klonis* and Ruben Pinedo-Cuenca
School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA, England

ABSTRACT

Enterprise Resource Planning (ERP) systems offer a certain degree of customisation, to accommodate different aspects of a business, but no ERP platform can be considered a panacea for every application. The selection of appropriate ERP software is a paramount factor for the effective execution of any implementation project. Whilst the selection criteria in large organisations have been based on actual organisational requirements, it has been observed that these are often overseen in small-medium enterprise over the software procurement costs. The initial capital expenditure savings may incur increased operational costs once the system is in use and other indirect costs such as people loosing confidence in using the system. The paper presents the findings of an ERP implementation project in the forging industry and discusses the challenges in tailoring the system to the business needs. The study is part of a larger improvement project being carried out in the company.

1. INTRODUCTION

The forging industry is one of the oldest metal processing industries [1]. This metal shaping technique changes a metal’s properties through heat treatment and hammering (hot forging) or through hammering at ambient temperature (cold forging). Whilst significant advances have been made in other metalworking industries such as casting and sheet fabrication, the forging industry still relies on rather old fashioned production methods and intensive manual labour activities. The old fashioned methods employed in the production systems are often reflected in the supporting pillars of process technology structures.

Developments in information processing technologies facilitated the computerisation of the set of procedures, data and knowledge used by the company to manage production and logistics to successfully deliver a product to the end customer [2] and also provided organisations with the opportunity to develop real-time linkages across the supply chain [3, 4, 5]. Enterprise Resource Planning (ERP) systems first appeared in the early 1990s and initially provided inventory control [6]. They were then expanded to include MRP and MRP II, and finally integrated all departments and functions of an enterprise such as sales and order management, marketing, purchasing, warehouse management, financial accounting and human resource management into a single business management system [7].

The benefits of ERP systems have been well documented [8, 9] as are the difficulties related to their implementation [10, 11, 12]. The benefits of ERP systems in large organisations, in addition to the lessons learned have made ERP plausible for small-medium enterprises [13]. These have previously been reluctant in investing in IT technologies [14] due to the complexity, duration and cost implications associated with IT implementations [15].

The implementation of ERP systems can be a difficult task [16] which can fail in spite of the large amount of capital and time invested [17]. The selection process of suitable ERP solution has been identified as a crucial task in the implementation process [18]. In the above studies, the authors acknowledge that one of the primary concerns for SMEs is cost, however they argue that the selection of the ERP software should not be made just on the basis of cost, as the operational implications could have a greater impact. Another study [19] explored the differences in the selection process in SMEs and large organisations and concluded that the priorities of the organisation in respect to ERP selection in addition to the selection process costs differ on the basis of size. The authors suggested that the selection criteria for large organisations included increasing organisational flexibility and enhancing supply chain relationships with customers and suppliers whereas for small-medium organisations, the criteria included cost and adaptability of the software.

* Corresponding author: Tel.: +44 1642 342482; Fax: +44 1642 342401; E-mail: t.klonis@tees.ac.uk
In the absence of adequate empirical research in the implementation of ERP software in the forging industry, this study provides an overview of the implementation process, highlighting the issues and pitfalls for future adopters. The results of this case study suggest that cost is indeed the primary selection criterion for SMEs and present the operational challenges associated with the implementation of ERP software, selected purely on the basis of capital expenditure.

2. BACKGROUND

The company of this case study is a UK manufacturing SME specialising in the supply of bespoke components for critical engineering applications. The company’s headquarters and manufacturing facility are based in Scotland, whilst the forging plant is based in the North East of England. The forging facility is currently the home of 45 employees and has been operating since the 1970s, manufacturing chains and lifting tackle. In the early 80s the plant started manufacturing goods for the UK and Norwegian oil and gas industry and in the late 90s expanded its production facilities to handle large forging work that was previously outsourced.

The forging plant installed its first management information system, built on MS DOS, in the late 80s. According to the general manager, the system was state of the arts at the time, and included modules for handling accounts, creating sales orders and production route cards and controlling stock but lacked important capabilities such as estimating and production modules. Activities such as creating estimates, raising quotes, quality and dispatch documents were carried out using external applications such as departmental databases, spreadsheets and word documents. Albeit employees were rather comfortable using the DOS system, the latter failed to communicate information adequately across the departments and did not provide real time stock visibility, causing major delays and problems in the business. As a result duplicate data entry was a common occurrence across all departments taking as much as 50% of the orders lead time. In spite of the laborious information checks from the quote stage to the production stage the quality of entered information was questionable. This had unwanted effects on two accounts: first it jeopardized the quality of information handled by the business, and second it increased the order processing time before the creation of a route card, consuming valuable lead time.

Over the last twenty years the business was temperamental due to the uncertainty in the oil and gas market. Recently however, the company was given the opportunity to grow significantly from £ 3.5 million turnover to £ 8 million, as it was designated as the preferred supplier for a large multinational organisation. The subsequent increase in sales volumes raised the amount of paperwork exponentially, to the degree where valuable time was lost due to the inefficiency of the legacy system. The company had already installed an ERP system in its manufacturing facility in Scotland; which albeit not being used to its full capabilities had resulted in significant reductions in data entry and processing time. Following this implementation, management moved with the purchase of additional licences to replace the aged DOS management information system in its forging facility. Unfortunately the implementation of the system in the forging facility was not as straight forward as initially thought, and after two years of purchasing the additional licences, the company was still using the legacy system. With plans to grow substantially by 2011 and in an attempt to effectively accommodate the rapid and demanding changes the company liaised with Teesside Manufacturing Centre (TMC) to:

- Implement the ERP software.
- Introduce a continuous improvement culture within the production workforce
- Reduce lead times for handling orders and generating route cards
- Introduce adequate stock and quality control mechanisms.
- Eliminate duplicate data entry
- Streamline laborious processes to control specialized customer requirements.

3. TIE-CHI METHODOLOGY

The implementation of any improvement initiative has been reported to yield changes in the organization, which are viewed reluctantly and resistively [22]. The project was carried out applying a proof methodology developed by TMC and well-recognised in the North-East of England. The methodology, known as Total Integrated Enterprise (TIE-CHI) [20, 21] was developed to overcome the main implementation challenges regarding management and cultural change. Different from traditional approaches, the TIE-CHI methodology uses a dual approach to ensure
from the outset of the project, the top management support and commitment, and the ownership of the project across middle-management and workforce. The methodology entails three stages:

- **Consensus**: to create consensus of the needs and priorities of the business across the management layers of the organisation,
- **Homing-in**: to define strategic, tactical, operational, technological and cultural needs and gaps,
- **Implementation**: to embed the selected initiatives in the organisation.

Upon completion of the Consensus stage, the management team was made aware about the need to enhance business practices and systems in order to remain competitive in the current economic environment. Business needs and priorities were defined and communicated to the rest of the organization in order to condense the resistance to change by getting people motivated and involved with the implementation [23]. After the whole organization was aware of the importance of the project, the Homing-in stage was initiated, where workforce and medium management was involved to define the needs and gaps of the organization from the social, cultural, technological, operational and strategic point of view. The authors held an introductory session highlighting the unique opportunity for the implementation team to re-engineer their business processes by proposing new and innovative ideas, incorporating best practices based on the actual company needs, rather than trying to mimic current practices. Bearing in mind the company’s constraints and company priorities and needs, the implementation was divided into two phases. Phase 1 included enquiry and sales process, engineering, quality, production planning and stock control. Phase 2, which is still ongoing, included advanced manufacturing practices, business performance measurement and integration with lean practices.

Process maps captured during the homing-in stage were fleshed and analysed to define gaps and limitations of systems [24]. During this process, participants were challenged to re-engineer current practices in paper by emphasising weakness in their practices and reflecting upon the need for change using the 5-Whys approach. At the end of the exercise, participants were able to perceive how their processes could contribute to achieve the business objectives and fulfilling the needs of the business. Each participant was able to experience how the system will overcome limitations with current practices and fulfilling the initial expectations. However, there were occasions where the system could not satisfy the initial expectations regarding process, data, and analysis and post-it notes where attached to the step in question as shown in Figure 1.

![Figure 1: Process Maps](image)

The process maps were further developed using process activity mapping and the lean principles of eliminating waste or non-value adding activities. The application of lean principles and the cultivation of a continuous improvement environment have been used successfully in other studies see for example [25, 26]. The process activity mapping highlighted the opportunities for improvement and the areas that the new system would replace. This approach assisted the team to visualise the areas required for change and areas directly or indirectly affected, but overall, smoothing the negotiation process when the adoption of new process involved the introduction of new tasks to other departments.
At the enquiry and sales stage, the company received an average of 300 enquiries a month, at an order conversion rate of 25%. The company’s target was to generate a quote within 48 hours from receiving the enquiry; however this was hardly accomplished, as over 32 enquiries had to be handled daily. Upon receiving an enquiry, an enquiry number was manually recorded at a logbook. A hard copy of the enquiry was passed onto the estimator who used an access database to generate the quote, which was then communicated to the customer. If the quote was successful, the same information was re-entered into the DOS system to enable the creation of the Work Order. The process was cumbersome with multiple duplicate data entry stages and failed to provide adequate traceability relative to the order. In addition, there was no system in place to record the reasons behind lost enquiries, and the company failed to chase lost enquiries which had an impact on sales.

All accepted sales orders were passed on to engineering for the work order route cards to be created. These were generated using the DOS system. Different route templates were available for the different grades of materials and for different material processing specifications, enabling the user to select and create the route card in a straightforward manner. The DOS system listed the quantity of material available in stock, but failed to provide visibility of any material being allocated to a particular job. As such, there were instances where route cards reached the shop floor listing material reserved for other jobs, creating upset and delays in the processing of the orders.

The delays in processing route cards were prolonged further due to a significant amount of time spent at the quality stage to ensure all information on the route cards was correct and caused an adverse effect on quality. Any changes required were marked on the route card hard copy and passed back to engineering. If no changes were required, the Route Card was passed on to Production to be planned. Quality was the most important part of the business as all products had to be made according to rigorous specifications and all material had to comply with the exact customer specifications. This required an immense amount of paperwork which was created by a combination of word documents and outputs from the DOS system.

Production planning was another grey area within the company, as it simply followed the first in first out rule. Every job passing through the quality stage was entered into the production plan and was driven by the required despatch date. There was no indication of actual lead times and monitoring of WIP relied on shop floor employees who used a series of spreadsheets to record the completed stages of each job. In effect, the process of updating the production plan was time consuming and involved a significant amount of manual data entry. Without any means of prioritising jobs, orders were often late which affected the delivery performance.

4. IMPLEMENTATION CHALLENGES

It is widely accepted that ERP implementations can be lengthy and complex [27, 28, 29]. The main obstacles during EPR implementations fall broadly in four categories; people cooperation and commitment, cultural change, systems capabilities and knowledge base and learning process.

In this study people cooperation and commitment was not really an issue, as all employees did their best to fulfil their everyday tasks and their roles in the implementation. However, the load of the daily activities of some employees did not allow them to participate as much in the implementation, impacting on the project plan and creating conflicts of interests with other departments that had the opportunity to progress faster with the implementation.

Cultural change is a difficult process to manage, as it may be obstructed by a large number of factors. In this study, cultural change was threatened mainly by people’s fears of becoming redundant as their daily tasks could be significantly reduced or even eliminated; losing control of their functional areas or not being able to hide inefficiencies. These factors were overcome with the assistance of top management which ensured their positions were not at risk and by educating people in best practices and continuous demonstrations of how the system would mimic their systems in a more effective manner, enabling them to be proactive rather than reactive.

The system capabilities has an important role in the implementation process, due to the complexity added to the re-engineering process to adapt the business processes to the system or otherwise, to customise the system accordingly to the business as illustrated in Figure 3. If the software can not handle aspects necessary for the success of the business, then this will have a significant impact on the project costs, duration and expectations point of view, unless project risks are defined and mitigated. The ERP software in this study was chosen on the improvements realised in a precision machining environment, which is largely differentiated from a forging environment. It was soon evident that the system presented significant limitations to cope with some of the key priorities and needs for the business. In a precision machining environment, part costs are easily calculated by considering material costs, sub contract costs, setup and run costs, tooling costs and labour hours. The operations and machine utilisation are
more closely monitored and more in depth visibility may be obtained. The forging industry dates back many decades and traditionally the cost associated to forging a job is rather arbitrary calculated by taking into account the shape difficulty factor and the “material difficulty factor”, or forgeability (the ability of the material to deform without failure regardless of the pressure and the load applied) [30] [31].

Figure 2: Adaptation strategy

In an effort to reduce capital expenditure, the company had signed an agreement with its raw material supplier, for the provision of consignment stock. The company was supplied raw material billets at regular intervals and was charged for the amount of material that had been consumed each month. The ERP solution did not distinguish between standard stock and consignment stock. Therefore, the raw material cost was input when the material was booked in and not when the raw material was used. As a result, the system provided inaccurate financial information including material expenditure, stock value, job margins, invoicing, etc when consignment stock was used.

The importance of quality in the nature of business has already been highlighted. Full traceability from raw material to end product was a paramount quality requirement, as were the non conformance reports (NCR). The ERP solution did not have any functionality for inputting the raw material composition, used to identify suitable material of a particular composition. In addition, the ERP solution did not include a sophisticated quality module to handle the activities carried out by the quality control department such as recording mechanical test results and inspection results. The user was able to create NC reports; however this functionality did not restrict the user to issue material to next operations, harbouring the risk of moving unapproved material that has failed an inspection stage through the remaining production stages.

The vendor committed to develop part of these functionalities within an agreed timescale free of charge, however these had to be aligned with the vendor’s strategy for software development. Although this was a major risk for the overall implementation as all delays in delivering the promised functionality could impact on costs associated with the development of reports, the vendor delivered some of the required functionality enabling the team to create a solid solution definition for the company. The stock issue was resolved and the system was able to hold all the necessary information relative to the material composition, enabling the users to record the majority of the test results. However, certain aspects of the business such as estimating and creation of test certificates had to be handled by external applications and reporting mechanisms until the software vendor was able to incorporate these changes in the system.

Knowledge base and learning process can also impact the implementation process. In this occasion, age, limited IT skills and lack of theoretical background in ERP concepts where the main obstacles that affected the performance and speed of the implementation. Key people in the process found difficulties to retain the processes or understand the need from a strategic perspective. This was overcome by using a deep-learning approach, supported with special
training material designed to support learners. The use of e-learning material contributed to smooth the familiarisation of the user with the system.

5. PILOT AND PREVENTATIVE TROUBLE SHOOTING

Going live is a major challenge for any organisation. Solutions definitions generally follow the 80/20 rule, leaving 20% of processes untested as being perceived as sectional cases. However, the lack of understanding of the implications of these processes can jeopardise the efficiency of the system and distort information crucial for the management of the business. To minimise this risk and ensure that no processes had been neglected, the system was set according to the Solution Definition and employees were asked to use the system in parallel to their current system. The team decided to pilot a large order from their main customer. The team used the system to log the enquiry, create the quote, raise the sales and work orders and issue the WO to the shop floor. The pilot project highlighted some minor concerns previously missed during the training sessions; however they did not result in any major changes in the design of the process. After resolving the issues highlighted in the pilot, the company went “live” with the new ERP solution nine months after the start of the project.

6. IMPLEMENTATION BENEFITS

The potential benefits derived from ERP systems have been well documented in literature. These can differ from company to company as they are directly related to the company’s attitude to the project and the use of a sound methodology and skills to drive the implementation. To date, the company has been able to realise significant benefits through the ERP implementation. These have been distinguished in relation to the process studied (Table 1) and in relation to the business as whole in Table 2.

Table 1: Process tangible benefits (Phase 1)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Current State</th>
<th>ERP Future state</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead time</td>
<td>10.19 days</td>
<td>3.35 hrs</td>
<td>95.3%</td>
</tr>
<tr>
<td>Process time</td>
<td>11.3 hrs</td>
<td>38 min</td>
<td>60.47%</td>
</tr>
<tr>
<td>Delay time</td>
<td>8.5 days</td>
<td>1 hr</td>
<td>98.32%</td>
</tr>
<tr>
<td>No of ops</td>
<td>24</td>
<td>12</td>
<td>50%</td>
</tr>
<tr>
<td>Processing cost</td>
<td>£480</td>
<td>£280</td>
<td>41.66%</td>
</tr>
</tbody>
</table>

Table 1: Business-wide tangible and intangible benefits (Phase 1)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Current State</th>
<th>ERP Future state</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Accuracy</td>
<td>42%</td>
<td>80%</td>
<td>38%</td>
</tr>
<tr>
<td>Data accessibility</td>
<td>17 min</td>
<td>2.5 min</td>
<td>14.5 min</td>
</tr>
<tr>
<td>Quality Errors due to</td>
<td>25%</td>
<td>2%</td>
<td>23%</td>
</tr>
<tr>
<td>inaccurate data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Update information across</td>
<td>5 hours</td>
<td>1.5 hours</td>
<td>3.5 hours</td>
</tr>
<tr>
<td>departments (per day)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity planning</td>
<td>None</td>
<td>Medium</td>
<td>50%</td>
</tr>
<tr>
<td>Traceability</td>
<td>Low</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td>Business analysis</td>
<td>Low</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td>Business understanding</td>
<td>Low</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td>Business focus</td>
<td>Low</td>
<td>Medium</td>
<td>25%</td>
</tr>
<tr>
<td>Integration with parent</td>
<td>None</td>
<td>Excellent</td>
<td>100%</td>
</tr>
<tr>
<td>company</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Departmental</td>
<td>Low</td>
<td>High</td>
<td>50%</td>
</tr>
<tr>
<td>Communication</td>
<td>Medium</td>
<td>High</td>
<td>25%</td>
</tr>
</tbody>
</table>
7. SUMMARY

The paper presented an ERP implementation case study in the forging industry using the TIE-CHI methodology. The key steps undertaken in the project were outlined, highlighting the adaptation process and the implementation challenges and their implications relative to time and costs. The results of the study support the argument of other studies [32, 33, 34] and suggest that a rigorous selection process should be undertaken prior to the purchase of an ERP platform to avoid the risk of re-investing in an ERP solution. ERP systems are not designed for every application, and what works well in one environment may not work in another.

A working solution was delivered at the end of the study and the company went live with its new ERP system, realising significant benefits, both tangible and intangible. The intangible benefits of the organisation, were monitored at different stages of the project in order to identify changes in people’s perception about the organisation and the project. From the comparison, between the stage before to start the project and after the going live, the company recognised having changed to some extent their attitude regarding taking ownership of their processes to improve them and assist other functional areas. Furthermore, the authors also noticed significant changes in people’s involvement, changing from being spoon-fed and reluctant to use the system, to becoming very demanding regarding using additional functionality of the system, extracting management information from the system and monitoring key activities of the business. In addition, workforce seemed to work using a more holistic and integrated approach, feeding information to other departments as data concerns were raised.

ACKNOWLEDGEMENTS

The authors would like to thank One North East for partially funding this project and all the employees for their help and cooperation during this study.

REFERENCES


Product/process design and optimization is typically aimed at manufacturing a single product for a single customer, and often leads to underutilization of the available production capacity. Thus it is reasonable for a manufacturer to make an effort to minimize or eliminate that excess capacity to achieve high effectiveness. Excess capacity can be allocated by running multiple products/processes design and optimization independently of the first. However, exploring possible synergies between the two (or more) products/processes may bring even higher benefits.

This paper presents a case where a manufacturing process (plastic blow molding) was shared between two different products for two different customers with two different sets of needs. These customer needs were mapped into core value-creating processes, recognizing both the differences in their requirements as well as the similarities in their expectations. Conflicting differences in production volumes and quality requirements were reconciled leading to higher customer satisfaction and improved cost performance.

That was achieved by using an innovative approach in applying Quality Function Deployment (QFD) simultaneously to both sets of customer requirements, analysis of synergies and conflicts within each and also between the two. The goal was to find the common operating range to fulfill entirely both sets of customer expectations, under assumption that an overlap existed.

The effort focused on processes for two products: C1 and C2, where the challenges included: (i) strong process coupling (over 60%), (ii) product differences (in terms of feature complexity, production volumes, cycle times, and quality requirements). Application of the proposed approach resulted in a revised production line layout, which led to reduced scrap rates (below 1%) and increased production rate (by 24%).

1. INTRODUCTION

Quality Function Deployment (QFD) has been used extensively in the last few decades to generate a bridge between the Voice of the Customer (VoC) and Process Parameters, [1, 8, 9]. The two extremes of Product Design seem so far apart that many times, by the time we finally arrive to defining the Process Parameters there is no link between them. The journey from one end to the other includes: VOC, Critical to Satisfaction (CTS) elements, Functional Requirements (FR), Design Parameters (DP) and Process Parameters (PP). By using QFD, many successful companies reduced dramatically the Product Development cycle while increasing the Customers’ satisfaction, [6, 7].

This paper attempts to go beyond the common use of QFD methodology. It is rather rare for a company to have only one customer or to have the resources to deal independently with multiple customers. Nowadays, in order to survive, the companies need to continuously optimize the use of available resources while maintaining a high quality standard. We attempt to develop a methodology that would explore QFD and other concepts in resolving conflicts between multiple customers that have different needs to be addressed simultaneously in order for the supplier to be competitive, [6, 8]. At a minimum, the paper will provide a methodology that would allow companies to make the right decision when faced with the challenge of coupling or decoupling the simultaneous fulfilment of the needs of different customers.
For simplicity, the paper addresses the case of three different entities that compete for the same pool of resources: External Customer 1 (C1), External Customer 2 (C2) and the Internal Customer (C0). To add complexity, C2 requires two different product lines: C2-I4 & C2-V6. As mentioned in the Abstract, the needs of the two external customers are very different and they both differ from the needs of the C0.

The methodology will be elaborated upon a real life example that will prove what can happen if a proper analysis of the conflicting requirements is not performed in advance.

2. QUALITY FUNCTION DEPLOYMENT - QFD

2.1. Defining Customer Requirements and Critical to Satisfaction criteria

Quality Function Deployment, according to the definition in Dr. Akao’s book, [1], “provides specific methods for ensuring quality throughout each stage of the product development process, starting with design. In other words, this is a method for developing a design quality aimed at satisfying the consumer and then translating the consumers’ demands into design targets …”

As mentioned before, this paper attempts to optimize resources by using an innovative approach while meeting – or exceeding – multiple customer requirements. In Fig. 1 Mind Mapping was used to identify and cluster VoC. The paper deals with two external – C1 & C2 – and one internal – C0 – customers.

![Fig. 1 – CTSs deployment](image)

The “fuzzy” customer requirements were analysed and clustered in three categories: Cost, Quality and Delivery. In-depth analysis followed not only in identifying the Critical to Satisfaction criteria but also in benchmarking the plant performance against Best-In-Class. The first analysis is presented in Fig. 2.

Analysis of the 1st House of Quality – HOQ1 -:

- Customer requirements differ considerably between C1, C2 & C0. A short list of the major differences:
  - levels of complexity – C1 has 75% more components then C2 -;
  - production volumes – Vc2/Vc1 = 2 -;
  - cycle times - Tc1/Tc2 =1.5 -;
  - quality requirements – Qc1 > Qc2 -;
  - delivery – C1 expects JIT vs. C2 forecast based –

- There is a considerable gap in performance based on both, customers’ perception and internal performance;
- VoC elements on the left hand side of the house are listed in the order of importance; It can be observed that customer C1 is focused on quality versus customer C2 - on cost;
- In the “roof” of the HOQ1 we can observe the conflicts between CTSs; The major conflicts are between the # of Blow Molders (BM) & Assembly Lines on one hand and Days on Hand (Available Inventory of Finish Goods) & quality on the other hand.
2.2. Next step in our methodology is to analyze CTSs against FRs.

Some of the conclusions we can make based on HOQ2 – presented in Fig. 3 - are:
- The order of importance of FRs is very similar to CTSs and VoC s: Quality, Delivery and Economies of Scale; The challenge that this paper addresses goes beyond the traditional tripod balancing – Quality, Delivery & Cost – as each “conflicting” CTS has a different customer behind;
- Economy of scale conflicts once again – as seen in the “roof” of the house – with requirements related to Quality & Delivery

2.3. The last iteration of our QFD analysis is the analysis of the DPs against FRs.

Some important points in the analysis of HOQ3 shown in Fig. 4:
- The two most important DPs based on the previous analysis are: Two-way communication & Management System and Independent Value Streams (VS) – e.g. design independently the VS for C1 from the VS for C2
- In the “roof” of the HOQ3 – the contradiction area - we can clearly identify the main conflict between the independent Value Stream design and the requirements related to Cost - very important to both, C2 & C0;
- The final design needs to consider all the conflicts identified throughout the QFD development and either eliminate them or minimize the adverse effects.
3. DEVELOPING DESIGN CONCEPTS – PUGH MATRIX

Based on the QFD analysis, design concepts have been developed attempting to meet all the customers’ requirements. A summary of the designs – Fig. 4 – is followed by a Pugh Matrix analysis - Fig. 5 -.

The Datum against which all the proposed designs were evaluated consists of the initial design: Integrated BMs, coupled C1 & C2 and, within C2, coupled down-lines (DL) C2-I4 & C2-V6. The initial design was chosen without considering all the main elements required in order to meet the VoCs. That approach led to major conflicts throughout the manufacturing chain that could not be resolved.

Based on the Pugh Matrix analysis we can conclude the following:

- There is no “ideal” design as the three customers have very divergent requirements;
- 4 out of the 6 proposed designs are better options than the Datum;
- The 2nd best design – D1 – 100% Independent VS – meets most of the C1 & C2 requirements although it is unacceptable to C0 due to the very high cost associated;
- The best design – D3 – Independent BMs, independent DLs C1 & C2 and coupled C2-I4 & C2-V6 – is a compromise between the 3 opposite requirements;
4. APPLICATION OF THE METHODOLOGY IN A REAL LIFE INDUSTRIAL EXAMPLE

4.1. Case introduction and problem definition

The manufacturer of automotive products C₀ supplies directly to two different customers: C₁ & C₂. Due to targeting opposite sides of the market – luxury vs. high volume/low cost - the two customers have very different requirements – e.g. C₁ has the major focus on the quality of the product and timely delivery whereas C₂’s focus is first and foremost in cost reduction and diversity of the product – two versions: I4 & V6.

The design that was originally chosen is the Datum in our Pugh Matrix analysis: integrated BMs, coupled C₁ & C₂ and, within C₂, coupled DLs C₂-I4 & C₂-V6. That was mostly driven by the C₀ as the design met all the main requirements focused around economy of scales – e.g. using one BM for both, C₁ & C₂, one gantry system and one DL for both versions of C₂.

As shown in Fig. 6, the BM is shared by the two customers. On addition, due to the common gantry system designed to transfer the part to either one of the DLs but not to both simultaneously, the two customers could not build in parallel. Obviously, this design puts a lot of pressure not only in scheduling, but also in the way the lines are controlled and managed.

The design delivered very inconsistent results. Among other performance metrics, the scrap produced by the BM is presented in this paper. In Fig. 6 we can see in the embedded P-Chart that the average scrap rate was over 20%, [2]. That is an extremely high level comparing not only with outside competitors but also vs. other Value Streams in the plant. The delivery was also very poor requiring repeatedly expedited shipments to the customers due to the shortage of parts.

4.2. The solution

The situation required a return to the drawing board and a complete redesign of the production lines. A quick analysis was made and, based on it, a design where the BM is shared but the two external customers are completely decoupled was chosen. Going back to our QFD analysis, we have already identified that the best design, considering the constraints, is decoupling completely the two customers – including the BMs -. However, even if the chosen re-design was not the optimum, decoupling completely the assembly lines improved dramatically the business metrics.

In order to facilitate the decoupling, process optimization was initiated with the following goals shown in Fig 7:

- Reduce the travel of the Associates;
- Re-program the welder, a key piece of equipment in the assembly of the parts;
- Replace automatic scanners with manual;
- Introduce a new mistake-proofing station to reduce the rework for C₂;
- Re-balance the line so the manpower does not change with the introduction of the new station;
- Open a logical gate at the back of the feeding conveyors in order to improve the flow, [3, 4].

Fig. 6 – Initial Design with the Scrap P-Chart

Fig. 7 – Process Optimization on C2 – Section of Fig. 6

Fig. 8 – Constraints Reduction Chart

In Fig. 8 we can see the reduction in the design constraints by decoupling the DLs. The re-design is identical with Design 2 in our Pugh Matrix chart. We notice that the BM is still shared – that is against our best identified design – as well as the DLs for C2 – this time the decision is aligned to our best D3 solution.
4.3. The new layout

The major change in the new design stands in the decoupling of the two assembly lines: C₁ & C₂. Beyond the fundamental improvements identified by the Pugh Matrix, additional benefits turned a very problematic manufacturing process into an acceptable one that carried on the production until customer C₂ discontinued its product. Some of the technical and management improvements are listed below:

- A logical gate was designed & installed at the end of the input conveyor after the BM –Fig. 9--; That allowed a better management of the production – e.g. in case of assembly issues down the line the shells were recuperated and introduced back once the assembly line issues were solved; An expansion/contraction study was required in order to validate the reuse of the unfinished product;
- Customer C₁ was relocated in such a way that access from both sides was provided to the Associates; That solved many of the production issues, allowing a more efficient distribution of work as well as easier access to replenish the inventory, [5];
- Incremental steps were taken towards improving both assembly lines from the flow as well as safety & ergonomics perspective;
- One of the major improvements that allowed a much better scheduling based on customer demand was change over reduction on the BM;

Fig. 9 – New Layout

4.4. Improved metrics

In Fig. 10 we can see the dramatic improvement in the scrap level generated by the BM from over 20% to below 1%, [2]. All the improvements mentioned above had contributions to the improved performance.

In Fig. 11 we can clearly see the positive trend throughout the improvement cycle. Also, although definitely correlated with the reduced amount of scrap, a significant improvement occurred in the number of parts produced from an average of about 525/day to over 650/day – 24% increase -. There is still a relatively high variation in the output that would definitely be addressed by choosing the design with the highest score in the Pugh Matrix analysis: D3 – independent BMs -.
5. SUMMARY AND FUTURE WORK

This paper developed a methodology that allows a company to resolve conflicts between multiple customers that have a rather wide range of needs. That is solved by using concurrently and innovatively concepts used traditionally either in Process or Design Optimization. The methodology was applied on a real manufacturing process faced with the above mentioned challenges. While an ideal solution might not be always possible, by applying the concepts presented in this paper companies could resolve either entirely or partially the conflicts generated by multiple customers requiring access to a limited pool of resources.

A logical continuation of this type of effort would be to verify the usefulness and applicability of the methodology in other fields – e.g. healthcare, food industry, academia, retail etc. –. The researchers plan also on expended the methodology by engaging business concepts – e.g. Evidence Based Management, Advanced Statistics etc. –. Considering that the attempt will be successful, the methodology will allow companies to make the right decision – from all three perspectives: Business, Design & Process - when faced with the challenge of coupling or decoupling the simultaneous fulfilment of the needs of different customers.

REFERENCES (Selective)

CAD of Main Spindle and Feed Drives for CNC Machine Tools

Zoran Pandilov* and Vladimir Dukovski

Department of Production Engineering,
Faculty of Mechanical Engineering, University "Sv.Kiril i Metodij"-Skopje
MK-1000 Skopje, Republic of Macedonia

ABSTRACT

Characteristics of main spindle and feed drives highly depend upon skillfulness of composing variable speed motors (AC or DC) and mechanical transmission elements. In order to enable interactive computer aided design and analysis of different design variants of CNC machine tool drives, original computer programs have been developed.

1. INTRODUCTION

This paper presents our investigation of principles for analysis and design of CNC machine drives. Contemporary development of machine tools highly depends on improvements in drive systems.

A special feature of CNC machine tool drives is use of variable speed (AC or DC) motors which provide continuous changing of cutting speed and feed rate.

The use of variable speed motors creates a question of their appropriate composing with mechanical transmission elements in order to get better output characteristics and to satisfy some conditions as a system.

2. THEORETICAL CONSIDERATIONS AND COMPUTER PROGRAM FOR DESIGNING MAIN SPINDLE DRIVES FOR CNC MACHINE TOOLS

Main spindle drives for CNC machine tools must provide constant power at wide range of speeds on the output of the main spindle. They consist of three parts: 1. variable speed motor, 2. mechanical transmission elements which provide appropriate output characteristics of the main spindle and 3. main spindle.

Usually mechanical transmission elements consist of: belt transmission or combination of belt transmission with gearbox (with two, three or four speeds).

Intensive development of quality tool materials enable using of very high cutting speeds and power.

Necessary output power on the main spindle can be calculated as:

\[ P = \frac{F_t \cdot v}{60 \cdot 10^3} \text{ [kW]} \]  

where: \( F_t \) - tangential cutting force component [N]; \( v \) - cutting speed [m/min].

CNC machine tools are used for production of workpieces with different shapes, dimensions and materials, with wide range of cutting data.

For ensuring these requirements the speeds on the main spindle must be regulated in very wide range,

---

* Corresponding author: Tel.: (389 2) 3099-259; Fax: (389 2) 3099-289; E-mail: panzo@mf.edu.mk
\[
R_{ms} = \frac{n_{max}}{n_{min}} \cdot \frac{v_{max}}{v_{min}} \cdot \frac{D_{max}}{D_{min}} = R_v \cdot R_d
\]

where: \(R_{ms}\) - range of regulation of output main spindle speeds; \(R_v\) - range of regulation of cutting speeds; \(R_d\) - range of diameters of the parts, or of the cutting tools; \(n_{max}\), \(n_{min}\) - maximal and minimal main spindle speed; \(v_{max}\), \(v_{min}\) - maximal and minimal cutting speed; \(D_{max}\), \(D_{min}\) - maximal and minimal diameters of the parts or the cutting tools.

According to our empirical investigation the range of regulation of main spindle speeds for CNC machine tools usually is within \(R_{ms}=20-350\) (exclusively rare to 600). Such kind of wide regulation of main spindle speeds needs particular attention in selection of variable speed motors and mechanical transmission elements.

Fig. 1: Power-speed diagram of variable speed AC motor

Fig. 1 presents power-speed diagram of variable speed motor, where \(n_{min}, n_{nom}, n_{max}\) are minimal, nominal and maximal speed of the motor, and \(P_m\) is nominal power of the variable speed motor.

Usually range of regulation of speed at constant power of variable speed motors is (2-8) (sometimes reaches values 12-16), which is far below required range \(R_{ms}=20-350\).

The overall range of regulation of output main spindle speeds can be calculated as:

\[
R_{ms} = R_{msm} \cdot R_{msp}
\]

where: \(R_{msm}=2-50\) - range of regulation of main spindle speeds at constant torque; \(R_{msp}=2-45\) (exclusively rare 70) - range of regulation of main spindle speeds at constant power.

There are two alternative methods of obtaining wide range of main spindle speeds at constant power: overrating of the AC or DC motor or combining the motor with gearbox with two, three or four speeds.

The second solution with two, three or four speed gearbox is widely used at the CNC machine tools.

Selecting the number of steps \(Z\) of the gearbox is in the range of regulation of the variable speed motor with constant power \(R_{mp}\), while with using the range of variable speed motor with constant torque \(R_{mm}=R_{msm}\) the whole range of regulation of output speeds of the main spindle is obtained [9,7,6].

Because of that, we can write:

\[
R_{msp} = R_{mp} \cdot R_z
\]

where: \(R_{mp}\) - range of regulation of variable speed motor with constant power; \(R_z\) - range of regulation of the gearbox. Variable speed motor can be treated as a particular group of gearbox with continuous changing speeds, which is first in the kinematic chain, with infinitely large number of transmissions, with transmission ratios which obtain geometrical progression with progression ratio \(\varphi \rightarrow 1\) and range \(R_{mp}\).
Gearbox can be treated as a transmission group which extend the speed range of the motor at constant power. Because of that characteristic of a transmission group $\Phi_z$ is:

$$\Phi_z = R_{mp} \cdot \varphi$$  \hspace{1cm} (5)

Because $\varphi \to 1$, we obtain

$$\Phi_z = R_{mp}$$  \hspace{1cm} (6)

As,

$$R_z = \Phi_z^{(z-1)}$$  \hspace{1cm} (7)

we can write

$$R_z = R_{mp}^{(z-1)}$$  \hspace{1cm} (8)

With the substitution equation (8) in (4), we get

$$R_{msp} = R_{mp} \cdot R_{mp}^{(z-1)} = R_{mp}^z$$  \hspace{1cm} (9)

where: $Z$-number of speeds of the gearbox.

With known $R_{msp}$ and $R_{mp}$, using the equation (9), we can calculate the necessary number of speeds of the gearbox $Z_E$:

$$Z_E = \frac{\log R_{msp}}{\log R_{mp}}$$  \hspace{1cm} (10)

The equation (10) is recommended also in the literature [9,7,6].

Fig.2: Diagram P-n of the main spindle a) with $z=3$ and b) with $z=2$

Because $Z_E$ is usually a decimal number, it is round to the nearest full number.

If $Z > Z_E$ we get characteristic with overlapping speeds (fig.2a).

In case $Z < Z_E$ we get characteristic with step decrease of the power $\Delta P$ (fig.2b).

For example, if from equation (10) we get $Z_E=2.5$, than $Z$ can be 2 or 3.

In case of $Z=3$ we get P-n characteristic as in the fig. 2a, and if is accepted $Z=2$ we obtain characteristic as in the fig. 2b.
Percentual decrease of the power $\Delta P$ in relation with the nominal power $P_m$ of the motor, when $Z < Z_{E}$ can be calculated with the equation (11),

$$\frac{\Delta P}{P_m} = \left(1 - R_{m_p} \cdot \frac{Z_{E}}{Z} \right) \cdot \frac{R_{m_p}}{R_{m_{sp}}} \cdot 100 \%$$

(11)

Usually $\Delta P/P_m$ should not be greater than 30% [9,7,6].

The above algorithm is implemented in an original computer program written for PC in C language which enables interactive design and analysis of main spindle drives with different design characteristics.

3. THEORETICAL CONSIDERATIONS AND COMPUTER PROGRAM FOR DESIGNING FEED DRIVES FOR CNC MACHINE TOOLS

The feed drive consists of an electromotor and mechanical transmission elements. The mechanical transmission elements comprise all the machine parts which lie in the torque (power) transmission flow between the servo motor and the tool or workpiece. In different design variants the following mechanical transmission elements are most frequently used: clutches, ball lead screw and nut units, rack and pinion units, bearings, gears, gearboxes (planetary, cycloidal, harmonic), toothed belt gears, guideways etc.

The main task in the feed drive design is a selection of a servo motor and mechanical transmission components. During this process the drive angular nominal frequency $\omega_{od}$ and nominal angular frequency of the mechanical transmission elements $\omega_{omech}$ are calculated.

In order not to affect the properties of the highly dynamic AC or DC servo motor, the nominal angular frequency of the mechanical transmission $\omega_{omech}$ elements must be higher than the drive nominal angular frequency $\omega_{od}$.

According to [1,10,6]

$$\frac{\omega_{omech}}{\omega_{od}} \geq 2$$

(12)

is recommended.

To satisfy the requirements and to enable a long exploitation period particular attention has to be paid to the selection of feed drive servomotors. An improper servo motor selection results in a less efficient operation of machine tool and a short exploitation period.

Total load torque $M_{tot}$ can be calculated as:

$$M_{tot} = M_{mf} + \sum M_{fl} \quad \text{[Nm]}$$

(13)

where: $M_{mf}$ is a torque caused by the machining force [Nm]; $\sum M_{fl}$ is a sum of torques caused by friction and losses [Nm].

The next step is a calculation of the necessary motor speed $n$ for a rapid feed rate.

The selection of a variable speed motor can be from a catalogue, or from an appropriate data base, developed during the investigation [6].

The total moment of inertia $J_{tot}$ can be calculated as:

$$J_{tot} = J_m + J_{ext} \quad \text{[kgm}^2\text{]}$$

(14)

where: $J_m$ is a motor moment of inertia [kgm$^2$]; $J_{ext}$ is an external moment of inertia reflected on motor shaft [kgm$^2$].

Equations necessary for calculation of $M_{tot}$, $n$ and $J_{tot}$ for different design variants are given in details in [1,6].
After calculation (Mtot and ne), for the selected servo motor an analysis of dynamic behavior must be performed.

With a dynamic behavior analysis, we calculate the acceleration time to rapid traverse feed rate for loaded motor ta, nominal angular frequency of the drive \( \omega_{od} \) and position loop gain \( K_v \).

The acceleration time to maximal speed for loaded motor \( ta \) can be calculated as:

\[
ta = \frac{J_m \cdot n_m}{9.55 \cdot M_a} \cdot 10^3 \quad \text{[ms] (15)}
\]

where: \( n_m \) is maximal motor speed \([\text{min}^{-1}]\); \( M_a \) is acceleration torque \([\text{Nm}]\).

The acceleration time to maximal speed of unloaded motor \( tb \) is:

\[
tb = \frac{J_m \cdot n_m}{9.55 \cdot M_a} \cdot 10^3 \quad \text{[ms]} (16)
\]

The acceleration time to the maximal speed of unloaded motor \( tb \) is given in a motor catalogue. If \( tb \) is not given directly, it can be calculated indirectly by the maximal angular acceleration of the motor shaft \( \alpha \) [rad/s^2].

Because

\[
M_a = J_m \cdot \alpha \quad \text{[Nm]} (17)
\]

equation (16) becomes

\[
tb = \frac{n_m}{9.55 \cdot \alpha} \cdot 10^3 \quad \text{[ms]} (18)
\]

With the substitution of equation (16) in (15)

\[
ta = tb \cdot \frac{(J_m + J_{ext})}{J_m} \quad \text{[ms]} (19)
\]

If \( ta \) is greater than a permitted value, corrections are made in mechanical transmission components (transmission ratio, feed screw lead etc.), in order to reduce \( ta \) and to satisfy the necessary value.

For an approximate mathematical calculation of nominal angular frequency of the drive \( \omega_{od} \), model shown in fig.3 [1], can be used.

\[
\omega_{od} = \frac{1}{T_{eld}} \left[ 1 + \frac{1}{2 \cdot \frac{T_{mech}}{T_{eld}}} \right] \quad \text{[s}^{-1}] (20)
\]
where: Teld is a drive electrical time constant [s]; Tmechd is a drive mechanical time constant [s].

Another important element which can be approximately calculated is the position loop gain $K_v$.

The position loop gain $K_v$ is a ratio of nominal speed $v_n$ [m/min] and difference between nominal and actual position $\Delta x$ [mm].

$$K_v = \frac{v_n}{\Delta x} \left[ \frac{m}{min} \right]$$

$$K_v = \frac{1000}{60} \frac{v_n}{\Delta x} \left[ s^{-1} \right]$$

The analysis in [1] shows that in ideal condition the optimal value of $K_v$ must lie in the range of:

$$0.2 \cdot \omega_{od} \leq K_v \leq 0.3 \cdot \omega_{od}$$

(23)

For real conditions it is recommended:

$$K_v < (0.2-0.3) \cdot \omega_{od}$$

(24)

The calculated values for $K_v$ from equations (23) and (24) are approximate. The exact value can be obtained experimentally by the numerical control of machine tool [6,8,2,3,4].

One of the most important requirements for good dynamic behavior of the feed drive is high acceleration of the CNC machine tool slide due to the demand for minimal mechanical time constant [1,5,6]. Magnitude of inertial forces which directly affect the accuracy depends on the magnitude of slide acceleration.

Acceleration limits are recommended [1,5,6]:

- for machine tools with normal accuracy ($a_{per} = 0.8-1.5 \text{ m/s}^2$),
- for machine tools with greater accuracy ($a_{per} = 0.2-0.4 \text{ m/s}^2$).

For the already selected type of servo motor, with corrections of some elements of mechanical transmission (transmission ratio, feed screw lead etc.), we may obtain a higher acceleration of the machine slide using the appropriate optimization procedure.

The acceleration of the machine slide is given as:

$$a = \frac{dv}{dt} \left[ \text{m/s}^2 \right]$$

(25)

For the variant with a ball feed screw and nut:

$$a = \alpha_1 \cdot \frac{h \cdot i}{2\pi} \left[ \text{m/s}^2 \right]$$

(26)

and for the rack and pinion variant:

$$a = \alpha_1 \cdot r_p \cdot i \left[ \text{m/s}^2 \right]$$

(27)

where: $v$ is a rapid traverse feed rate [m/min]; $h$ is a feed screw lead [m]; $r_p$ is a radius of the pinion [m]; $i$ is a transmission ratio; $\alpha_1$ is an angular acceleration of the loaded motor shaft $[\text{rad/s}^2]$. 
The angular acceleration of loaded motor shaft $\alpha_1$ is

$$\alpha_1 = \frac{Ma}{J_{tot}} \text{[rad/s}^2]$$

(28)

where: $Ma$ is an acceleration torque of the selected motor [Nm].

In that case equations (26) and (27) are transformed into:

$$a = \frac{Ma}{J_{tot}} \cdot \frac{h \cdot i}{2\pi} \text{[m/s}^2]$$

(29)

$$a = \frac{Ma}{J_{tot}} \cdot \frac{r_p \cdot i}{\pi} \text{[m/s}^2]$$

(30)

The optimization of a transmission ratio $i$, feed screw lead $h$ or radius of the rack pinion $r_p$ will be done by using the following procedure:

1. For every standard value of the feed screw lead $h$ or radius of the pinion $r_p$ we can calculate the transmission ratio range $i_1 \leq i \leq i_2$, to satisfy the following conditions:
   - the calculated necessary motor speed $n_e$ for the desired rapid traverse feed rate must be smaller or equal to the maximum motor speed $n_m$ ($n_e \leq n_m$),
   - the total load torque $M_{tot}$ must be smaller or equal to the nominal motor torque $M_n$ ($M_{tot} \leq M_n$).

2. In the range $[i_1, i_2]$ we will find the maximum of the function of acceleration $a = f(i)$ at constant $h$ or $r_p$:

$$\max[a(i)] = \max[a(i_1), a(i_2), a(e_1), \ldots, a(e_j)]$$

(31)

where $e_1, \ldots, e_j$ are extremes in the range $[i_1, i_2]$.

The extremes can be found by the equation

$$\frac{da(i)}{di} = 0$$

(32)

The function of the acceleration $a = f(i)$ in the range $[i_1, i_2]$ may have one, more or no extremes (fig.4).

![Possible forms of the function of acceleration](fig.png)

Fig.4: Possible forms of the function of acceleration

When the function of acceleration $a = f(i)$ gets a maximal value for the constant feed screw lead $h$ or radius of the pinion $r_p$, the transmission ratio obtains the relative optimal value $i_{opt}$.

3. To get the absolute optimum of the transmission ratio $i_{opt}$ and optimal values of the feed screw lead $h_{opt}$ or of the pinion radius $r_{p_{opt}}$ we should repeat $n$ times procedures described above in 1 and 2 for all standard values for $h$ and $r_p$. 
and $r_p$. In that way we will get $n$ relative optimal transmission ratios $i_{opi}$ for the appropriate $n$ different standard values for $h_i$ or $r_{pi}$, where $i=1,...,n$.

The pair $(i_{opi},h_i)$ or $(i_{opi},r_{pi})$ that gives the maximal value for the acceleration function, will provide the absolute optimum for the transmission ratio $i_{opt}$, and, the optimal value for the feed screw lead $h_{opt}$ or for the radius of the pinion $r_{opt}$.

It means

$$\max\{a(i_{opt},h)\} = \max\{a(i_{opt1},h_1),...,a(i_{optn},h_n)\} \quad (33)$$

or

$$\max\{a(i_{opt},r)\} = \max\{a(i_{opt1},r_{p1}),...,a(i_{optn},r_{pn})\} \quad (34)$$

Using equations (33) and (34) we obtain the pair $(i_{opt},h_{opt})$ or $(i_{opt},r_{opt})$ which provides a maximal value for the function $a=f(i)$.

This optimization procedure is different from procedures shown in [1,5], where the relative optimal transmission ratio $i_{opt}$ is calculated using equation (32) without taking in consideration $n_e \leq n_m$ and $M_{tot} \leq M_n$.

The theoretical assumptions treated in the text above, are implemented in the computer program, written for PC in C language.

4. SUMMARY

The created computer programs enable an efficient interactive and optimal design and analysis of different variants of CNC machine main spindle and feed drives. The presented software also reduces the design time and modernizes the design process.

REFERENCES


Interaction of natural inspired methods in process planning

Cornelia Teich¹, Daniel Kretz²*, Matthias Richter¹, Joerg Militzer⁴ and Tim Neumann⁵

¹ AWT Horwath
09130 Chemnitz
Saxony, Germany

²,³,⁴,⁵ Department of Economics
University of Applied Sciences Zwickau
08012 Zwickau, Saxony, Germany

ABSTRACT
A high response time to customer requests, plurality of product-variants and a rising pressure of prices are challenges facing businesses in the global-based economy. In particular, the use of integrated software solutions offers here the chance of a competitive advantage by reducing the processing time and increase flexibility in regard to product development. The integration of software solutions for the planning process - Computer Aided Design, Computer Aided Manufacturing, Computer Aided Process Planning - and production - Enterprise Resource Planning - has also in the future great potential to improve flexibility and lower the time to react to customer requests. Related to their specific domain these solutions provide good results. The combination and integration of automated process- and production-planning with a CAD Drawing is necessary, but not established until now. In the various planning stages good solutions are reached, but coherences are not well evaluated and implemented. If an optimized routing should be derived directly and automated from the design of the product you have to use a planning tool which connects the individual parts of computerized planning. Therefore a software is in development, which integrate the described planning tasks and additionally ensure the continuity of production, from design to manufacturing. For the optimization, nature inspired methods - such as Swarm Intelligence, Neural Networks in particular the Growing Neural gas, Genetic algorithms - are used to allow an integrated optimization of process planning. The learning ability of the individual solutions allows creating links between planning levels. The modules are: Generating of process variants, search for the resources needed, considering the utilization of resources identified, evaluation of internal and external interdependence. The aim of this paper is to present this model and include internal and external inter-dependencies in the integrated process- and production-planning software. The novel interaction of the natural inspired methods leads to addressing the topic.

1. INTRODUCTION
The use of integrated software solutions in process- and production-planning offers the chance of a competitive advantage by reducing processing time and increasing flexibility regarding product development. The integration of software solutions for the planning process, Computer Aided Design (CAD), Computer Aided Manufacturing (CAM), Computer Aided Process Planning (CAPP), and production where Enterprise Resource Planning (ERP) is used - has great potentials to improve flexibility and reducing the reaction time to customer requests in the branch of customer-individual manufacturing in the future.

In the various planning stages, good solutions are reached, e.g. CATIA for CAD and SAP for ERP-Systems. But there is no look for the coherences between the various stages. This is one fact for the deficit of these single solutions in the main task of process and production planning. The necessary combination and integration of process- and production planning to the design of a product in its implementation (number one in figure one) are not established until now. Additionally the use of natural inspired methods, such as Swarm Intelligence, Neural Networks in particular the Growing Neural Gas (GNG) or Genetic Algorithms (GA), to allow such an integrated optimization (number two in figure one) is not included.

* Corresponding author: Tel.: +49 (0) 375-5363265; E-mail: daniel.kretz@fh-zwickau.de
Two problem areas have been identified and worked on. First, develop an integrated planning environment for process-planning in product development especially using, second, a novel combination of natural inspired methods for producing the patency of process and production planning.

The starting point for developing the planning environment is a feature-based CAD-Drawing. With this drawing it will be possible to create automatically a working plan. This is performed in compliance with the main criteria of various target areas in the existing contracts of the company and its scheduling (Detailed Scheduling (DS)).

To reach the main targets or coherent targets of the company for example in DS, heuristics, specifically natural inspired methods (Swarm intelligence, neural networks, especially the Growing Neural gas, genetic algorithms) are used.

2. PRESENT SITUATION

Since the early 1950s, the investigations of scheduling problems have been in the focus of Operations Research (OR). However, it is difficult to use OR-results in real production environments. Reasons for this are as follows [13]:

- The many details of the recent application situation (alternate machines, cleaning times, setup time, etc.)
- the dynamic and uncertain structure of the production environment (random disturbances, unpredictable periods of preparation, etc.)
- in generally reluctant optimality criteria
- the need to interact with the user

To handle these and other multi-objective optimization-problems, it should be an advantage to use natural inspired methods or heuristics. A first example therefore is the method of Swarm Intelligence, with the basis of collective behaviour of decentralised systems. Simple agents are interacting between themselves and their environment. An important representative method is imitation of ants at foraging [4], but was initial developed in 1991 by Dorigo [3]. As further developments are the Particle Swarm Optimisation [8], the Fully Informed Particle Swarm [12] or the Rough Particle Swarm Optimization Algorithm [1] to mention. In the field of process-planning not many applications are known.

Another method to solve optimization problems are neural networks. They imitate natural learning models and systems, such as the human brain. The use of neural networks in an automated planning process is elicited and documented many times (e.g.[14] or [11]). Particularly, the Growing Neuronal Gas (GNG), initially described by Fritzke ([6], [7]), is distinguished by its capacity to add new nodes. These approaches consider individual sections of the automatic selection of mechanical resources. But there is no concrete examination of the situation in utilization of the company and no sequence optimization.

Newer approaches automate the planning process by suggesting automatically generated sequences and allocated the necessary resources. Khalifa et al. [9] describes, for example, a neural network, which allows determining the
available machines, tools and their capacitive situation. However, there is no integration of CAD drawings by an automated process planning. The inclusion and/or evaluation of different process variants are not established until now.

Genetic algorithms are heuristics, which imitate the natural evolution in reference to their population. In relation to combinatorial optimization problems in the production, numerous preparatory works are known. Tamura [15] for example uses a GA with heterogeneous population to solve multi-criteria planning-problems. Another Example is Fayad [5] who approximates to an unsure finishing date and those implementation with the aid of GA’s and fuzzy logics.

A simple method in the integrated production process is the repetition of planning (using saved work plans). Variant planning is very sophisticated. On the basis of input of certain work piece data, parameters for specific work plans are generated automatically and combined. In contrast to that procedure, the application of generative computerized work planning in practice is still fairly new. Basically there are four distinguished types of computerized process planning systems:

- variant planning systems
- generative process planning systems
- hybrid theories (combination of variant planning and generative process planning
- knowledge-based methods

The main task of production planning is to schedule processes in relation to the target criteria of the company. Recently, there are increased efforts to integrate the process and production planning and to prepare the continuity between CAD, CAPP, CAM and PPS. Amaitik et al. [2] developed a computerized environment for process planning with the integration of STEP (Standard for the exchange of product model data) AP224-Features. The use of features for process planning is dedicated to Tong et al. [16]. He describes how features are extracted from the neutral STEP AP214 data and implemented into a knowledge-based process planning. Interesting is the approach of Li et al. [10]. They assume flexible process plans which are optimized by genetic programming. This solution will be able to create suggestive process plans, but there is no look for the coherences to the definition in the drawing model.

In the various fields of research different problems are solved and their solution is documented. The integration of product- and process-planning has not been investigated until now. This leads to focus on a planning-environment which unites the various fields.

3. DESCRIPTION OF THE FRAMEWORK

One of the tasks is to study the interactions of different natural inspired methods on the basis of theoretical and practical steps, which should be clarified in the following chapters. Figure two gives an overview about the successive steps for the integration of process- and production-planning.
3.1. GENERATING PROCESS VARIANT PLANS USING SWARM INTELLIGENCE (FIRST STEP)

The basis of the developed software solutions are variants of the process, which are based on a CAD drawing. These variants are created by using natural inspired methods, e.g. swarm intelligence.

The problem of creating a process variant is similar to the travelling-salesman problem without returning to the starting point. All features of the CAD drawing must be attended to describe the product, but not necessarily in fixed order. Geometric and technological restrictions from the features on one hand and from the assigned process-steps on the other hand are necessary for the developed solution. Possible methods to generate the variant plans are for example Ant Colony Optimization or Particle Swarm Optimization.

One focus lies in the problem-specific mathematical formalization of relevant algorithms. As an opening solution, only a small number of useful technological variants are identified. These variants refer to the resources-equipment and their workload in the focused time-horizon. After steps two and three (see chapter 3.2 and 3.3), the target function can be determined.

Based on the best solution or the best solutions of the investigated algorithm with reference to the determined candidates, it is possible to interact with natural inspired methods to develop a certain amount of new solutions. These new solutions influence these steps again and can be assembled with others in process planning. This procedure is repeated and gradually leads to finding a certain (at least local) optimal sufficient solution.

The final output of the selected algorithm in this step is the provision of a technologically sensible, in reference to the target areas optimal variant of the process.

3.2. SEARCHING THE REQUIRED RESOURCES BY NEURAL NETWORKS (SECOND STEP)

After the generation of process variations for manufacturing, possible consequences of the various features of the CAD drawing (such as a sequence of intermediate products) have been created. Now the required resources have to be allocated. This happens on the edges of the process variant. These edges describe the essentially working step and are carriers of production time and costs.

For this assignment, in addition to the consideration of manufacturing features, the upstream and downstream states of the product (intermediate products) are relevant, because a matching of clamping device and work piece, weight, etc. must be made. In addition, the parameters of the features themselves influence the needed resources.
There are high-dimensional vectors on both sides, demand and supply, colliding each other. These vectors need a high dimensional search. Conventional search methods are not applicable here, because it is very difficult to point out the primary parameters. So, multidimensional index-structures are indispensable. A requirement of a feature in the specific context of the preceding intermediate product, the requirement vector, can therefore assign the appropriate resources. In the context of developing a custom procedure for determining the resources it will be an advantage to liberate this from Step 1 and model it as an artificial neural network.

By clustering similar resources it is possible to build competence bundles. A requirement of a feature in concrete context of the precedent intermediate product, the requirement-vector, can be assigned to the suitable resource. To develop a dynamic progress by adding new resources in the company, the GNG is preferred.

In this step it has to be referred to the interdependences which can lead to an increase or decrease of numbers of variants from the first step. For example, it can be possible that one feature will be made by two directly following producing-steps. Another possibility is that two or more features can be made in one producing-step. This case leads to an increase/decrease of features and thereby a increase/decrease of variants. Additional, this has the consequence that step one has to restart. These interdependences have to be described and finally build up in the learning-attitude of the neural network.

The result of this step is the second module of the software solution. The generated variants from step one and the required resources (from this step) are assigned to technological parameters that are necessary for step three.

3.3. Considering the utilization of the resources (third step)

The actual utilization of resources within the company finds the needed consideration in this task. If the company wants to get an assessment of the costs for the production of the component, it is not enough to use a quantitative-time viewing. Rather all existing contracts, a process variant as a result of step 2 and all the machines have to be considered.

This step ends in the valuation of a production variant in context with the utilisation of the company. The use of genetic algorithms proved to be right for machine scheduling and resource assignment in the last years. That is why GA’s are chosen for this step.

Because of the fact that the choices of the parameters and the kind of algorithm have an influence to the needed resources and the convergence of the method, it is necessary to accomplish comprehensive parameter studies. Furthermore, the primary target is to develop a target function of the optimization method. Additional to costs of material and production, it is essential to consider calculated costs for late delivery, non-compliance of price-agreement and image-damage.

From a mathematical point of view this is a multi-objective optimization problem in which the individual optimization goals compete in general, so they are often not complementary or indifferent. For this reason, in most circumstances candidates of solutions cannot be compared directly. Therefore it is necessary to develop a common objective function from the individual targets, e.g. using the AHP (Analytical Hierarchy Process) method.

Another result from this step is the opportunity to re-evaluate the first step with several new candidates of solutions. It is important to guarantee, that no solution will be dominated by another one. A disadvantage of these methods is that the targets should be clarified before running the AHP. But in some cases this will be very speculative for the decision maker.

The aim of the work in Step three is the third module of the software solution which evaluates the process variants with regard to the concrete capacitive situation in the company.
3.4. ELICITING OF INTERNAL AND EXTERNAL INTERDEPENDENCES (FOURTH STEP)

Due to the fact that a combination of these procedures for the integration of process and production planning is new, in this step the main target is to elicit the interdependences of the individual algorithms on practical application, some already described in the steps one to three.

The following interdependences are possible:

- Internal interdependences
  - Temporal dependencies of algorithms, especially for parallelization of the algorithms used in step one and three.
  - Change of the requirement due to structural dependencies of the logical design object.
  - Parametric dependencies of the various algorithms.

- External interdependences
  - Change of the CAD drawing
  - Change of resources structure in the company
  - Lack of resources and allocation of manufacturing steps outwards
  - Change of aim preferences in relation to costs, quality and so on

The goal of this step is to increase the flexibility of the software solution and to advance the predictability of the quality of the solution.

4. CONCLUSION

In this paper four steps are considered. Steps one to three have the aim to investigate, develop and combine the various methods for process and production planning.

For this the first module of the software, described in the first step, is already devised. The current task is to develop modular software which is used for the interaction of natural inspired methods in process planning, described in the second and the third step. This creates a planning system which elicits the interdependences in work plan. It creates an integrative planning environment which produces the continuity of product development.

Step four should elicit the dependences of the individual sections of the continuity software solution and implement obtained cognitions. Important results of the four steps are saved solutions which can derive for the next customer order. So they influence the following passes of the process.
REFERENCES


Computerized System for Progressive Die lay out Design

M.A. Farsi¹, B. Arezoo²

¹- Manufacturing Engineering Department - Islamic Azad University – Eghlid Branch, Iran
²- Manufacturing Engineering Group, Mechanical Engineering Department, Amirkabir University of Technology, Tehran, Iran

ABSTRACT

The sheet metal industry is very important today and sheet metal components are used in different fields such as Automobile, Aerospace and electricity industry. Application of the progressive die is one of the most important methods to produce sheet metal components, since it reduces production cost and time and increases quality of products. In progressive dies, two or more stations are used. In each station, one or more processes are applied. However, the design and manufacture of these dies are difficult. CAD/CAM systems have been proved to be very useful tools for this task. The main problem of CAD/CAM systems used in progressive die design is determining the operations sequence. In this paper, a system for determining the sequence of the bending and piercing operations is described. In this system, piercing and bending operation are separated and the sequencing of each type is done individually. For piercing stations, Fuzzy set theory is applied and the sequencing is done in one stage. For bending operations sequencing is done in two stages. First, the bending operations, which can be carried out simultaneously, are defined by a classification method. The sequence of the bending operation groups is then determined using fuzzy set theory. The capabilities of the system are shown using a sheet metal component.

1. INTRODUCTION

Progressive dies are used to transform flat strips of sheet metal into a formed part. This transformation is performed progressively by a series of stations that cut, punch, form, and bend the material into a desired shape. A progressive die that performs the various forming operations on the material is unique for every part. The various components that make up the die are located in guided and precision cut openings in plates, which are in turn located and guided by pins. The entire die is actuated by a mechanical press that moves the die up and down, and the sheet metal is fed through the die progressing from one station to the next. Progressive die design is a relatively complex and highly iterative process. Today's progressive die manufacturers face a number of issues that hold back improvements in productivity, quality, and turnaround time. These range from a lack of experienced progressive die designers, the difficulties involved in speeding up and improving quality in the more traditional design methods, and a shortage of any specific, easy-to-use, productive software solutions for progressive die design. In order to maintain their competitive edge and survive, more and more progressive die companies are looking to adopt three-dimensional computer aided design (3D CAD) technology and using computerized system for die progressive design.

2. RELATED WORKS

Researchers have tried for many years to develop a computerized system to design and manufacture progressive dies. In 1971, Shaffer [1] developed a Progressive Die Design by Computer (PDDC) system and Fogg and Jaimson [2] followed him by developing an improved PDDC by considering other factors which have an effect on die design. A weak point of this system is that it is semi-automated and the processing time is too long. Shibata and Kunimoto [3] developed a CAD/CAM system for which the aim is the screen-output of blank layout and die layout. Nakahara et al introduced a system of a progressive die design [4]. Choi et al. [5] developed a compact and practical CAD system for the blanking or piercing of irregular shaped sheet metal...
products and stator and rotor parts. Nee et al. [6] developed a computerized system to design and production of piercing progressive dies. Nee and et al. [7] used fuzzy set theory to determine bending sequencing. Kim et al. [8] developed a CAD system for the blanking and bending operation in sheet metal products. They used fuzzy logic in their system. They used 3 rules for piercing operations and 6 rules for bending operations sequencing. Farsi et al. [9] developed a new method to determine bending sequencing in progressive dies. Their algorithm was based on classification and fuzzy set theory.

These systems are mostly used as an assistant for the progressive die designer. Often, these systems have been focused on one operation type: bending or piercing operations. The number of the stations suggested by these systems is often more than the actual industrial parts. So the die layout suggested by these systems is usually modified manually by the designer. Also, the systems mostly recognize features interactively and feature recognizing is done using 2-D model or wire frame.

In this paper, a computerized system is described. This system is used for designing progressive die layout. This system includes feature extractor, manufacturability check and layout design modules. Features are recognized automatically from a solid model and the features are controlled using design for manufacturability considerations. At the end, using sequencing operations algorithms layout is designed and the sheet metal shape in each station is presented as 3-D model.

3. STRUCTURE OF THE SYSTEM

The system includes input and shape definition, Design For Manufacturability (DFM) and station layout design modules. Figure 1 shows the configuration of the system. This system developed in solidworks environment using Visual Basic (VB). The system recognizes sheet metal features and their data are saved in a file. These data are used to control production feasibility. Then features are separated to bending feature and piercing features. In the system first piercing operations and then bending operations are performed. The functional description of the system modules are presented in detail as follows:

![Figure 1. System configuration](image)

3.1. INPUT AND SHAPE DEFINITION MODULE

The features that are basic to sheet metal parts are stamped and formed features. Stamped features are categorized as internal and external features. Internal features include features like round holes, square holes, rectangular holes, oblong holes, etc. External stamped features include different types of notches. The most widely formed features are bends and flanges.

In the feature recognition module, all data used for the system are extracted. These data are used for DFM check, flating of the part and sequencing. Data modeling methodology in this work is based on object oriented approach. For example, the following data are extracted from a bend feature:

- Bend radius
- Bend angle
- Length of each bend
- Bend height
- Bending direction (up/down bending)
Coordination of start, end and middle points on bend line
Connected planes to bend
Bend factor (it is used for unfolding calculations)
Material thickness

3.2. DESIGN FOR MANUFACTURABILITY MODULE

The sheet metal industry is largely an empirical industry. The design and process planning of sheet metal products need human involvement more than other areas of manufacturing industry. Design of complex parts are based heavily on human experience. In the design stage, if some of the design considerations are ignored, the cost of production may increase rapidly. There for in this system a DFM module is provided and the system can control feasibility of sheet metal component by design considerations rules. These rules were saved in a data base. DFM rules were extracted from metal forming fundamentals, Industry observation, Hand Books, DFM books. All of the features recognized by input and shape definition module can be checked in this module. In the next paragraphs, some of the design rules are discussed. For more information see ref. [10].

a. Holes: The diameter of punched holes should be greater than the thickness of the work piece and not less than 2.5 mm. This is because the punch diameter becomes too small to bear the shear force required to punch the hole over a small area, often leading to failure.

Oblong holes: Oblong holes are generally incorporated as adjustment slots. As a general rule, the length to width ratio of an adjustment slot should be no more than 5 to 1.

b. Notches: Notches should extend inside the stock edge at least 1.5 times the thickness but not less than 0.50 mm. The minimum widths of tabs and slots should be 1.5 times the thickness or 0.5 mm. Their length should be a maximum of 5 times their width.

c. Bends: The width along the bend axis should be greater than or equal to 3 times the thickness. The bend radius \( R_b \), \( R_b \geq C \times t \) (t denotes the component thickness and C is a coefficient and depends on the material type).

d. Slot: The minimum widths of tabs and slots should be 1.5 times the thickness or 0.5 mm. Their length should be a maximum of 5 times their width.

3.3. STATION LAYOUT MODULE

This module is a computer aided process planning module to determine operations sequencing. Determination of operations sequence is the main problem in progressive die design. This problem is more important when the part has bends features. If \( N \) is the number of bending operations in a given part, then the domain of possible sequences in principle is \( 2N! \) [11]. Also, if \( M \) is the number of piercing operations, then the domain of possible sequences in principle is \( M! \). If the bending and piercing operations are performed separately then the total domain of possible sequences is \( M! + 2NN! \).

However, this number is usually limited due to geometrical and technical constraints. In other words, the number of possible sequences depends on the shape of the component. In this module, first the mother plane is determined. Then operations sequence is determined. At the end, the system presents the die stations layout as 2-D and 3-D model.

4. OPERATIONS SEQUENCING

Operations sequence is the most important and complex stage in a sheet metal component die design system. The system has to determine mother plane of the component and then determine sequencing of the operations using special algorithms.

4.1 MOTHER PLANE

Mother plane has a very important role in bending progressive die design. The mother plane is a fixed plane which stays without any rotations throughout the bending operations. All the rotating planes are called children planes. The rules for the determination of mother planes are as follows:

A plane surrounded by other planes
A plane located in the center of the part
The largest plane in the component
Figure 2 shows the mother plane for the sample part, the mother plane is colored in this figure. When the determination of the mother plane is not clear from the conditions as stated above, it is determined by the minimum number of bends between a plane and the plane in the central plane.

5. PROCESS PLANNING RULES FOR PIERCING OPERATIONS

The rules and the database of process variables are derived from plasticity theories, relevant references and the empirical know-how of experts in the blanking or piercing industries. The fuzzy set theory constructs a fuzzy matrix for calculating fuzzy relationship values and determines the optimum sequence by combining several rules with fuzzy reasoning.

Rule 1: Relation rule
Two types of relations among piercing features is-in and is-along are represented in the relation rule. Holes and slots in the plane are defined as the “is-in” relation; and notching, and piercing of the outer feature along the plane as the “is-along” relation. The fuzzy membership function of this rule is shown in Figure 3a.

Rule 2: Force and Shear length rule
One of the most important problems in the progressive die design is the forces and moment equilibrium. This affects the die lifetime and if the forces and moment are in equilibrium state then the lifetime is longer. The piercing force depends on the size of shear length of the shape to be pierced. Therefore, some of the die designers suggest which the piercing operation with larger force is performed in middle section of die. In other word, the smaller shapes are performed earlier. The fuzzy membership function of this rule is shown in Figure 3b.

Rule 3: Adjacent piercing rule
This rule is applied to adjacent shapes to be pierced so that the piercing is possible with a single punch shape, and the fuzzy membership function is shown in Figure 3c.

6. BENDING OPERATIONS SEQUENCING

In this system, bending sequencing is done in two stages. First, the bending operations, which can be carried out simultaneously, are defined by a classification method. In this method, all the bends are initially divided according to their bending directions (feed direction or perpendicular to it). Then for each direction, the bends are divided into operation groups according to classification rules. Three rules are used to determine the bending operation groups in this paper. These rules are based on relations between the bends in the component. The sequence of the bending operation groups is then determined using fuzzy set theory.

6.1 BENDING CLASSIFICATION METHOD
The classification technique is applied to the determination of simultaneous bends which can be performed in one station. The die designers use different rules to define simultaneous bends, because several parameters affect this procedure. According to experimental studies, many factors affect the determination of the simultaneous bends. However, the following rules can be summarized [9].

**Rule 1:** The bends that have bend lines along one line and whose bending directions are the same (up or down bending) can be performed in one station. Thus, they are said to be in one group.

**Rule 2:** The bends that have parallel bending lines and their bending directions are the same (up or down bending) and are placed on opposite sides of the mother plane can be performed in one station and are said to be in one group if the number of the planes between them and the mother plane are equal.

**Rule 3:** Related bends. In some sheet metal components, two planes can be related through geometric or dimensional tolerances. To obtain the tolerance and to comply with the positioning errors, these bends should be (or better be) performed together.

### 6.2 Fuzzy Set Theory

The handling rules and criteria needed for determining the bending sequence is discussed in this section. These rules deal with the selection of the next best bend for the bending operation. Each of these rules establishes relationships between pairs of bending operation groups. A high membership grade indicated for a particular rule means that the bend group is a good selection for the next operation according to this rule. These relationships between the bending and criteria are represented as fuzzy relations, and the membership grades of these fuzzy relations are determined through fuzzy membership function, as shown in Figure 4.

Fuzzy relations or sequencing rules describe the priority of each group. Thus, definition of these rules for a computer aided system is important. According to previous studies and experience of the authors, the following rules are suggested:

**Rule 1:** Distance rule. This rule describes the influence of the shape of a bend on the sequencing strategy. The further a bend is away from the mother plane, the higher its grade will be and thus should be bent earlier. The fuzzy functions presented in Figure 4a are used to determine the grade of membership by this rule.

**Rule 2:** Number of bends in a group. The higher the number of bends in a group, the more impact it will have on the overall shape of the part and vice versa. The more impact a group will have, the later it should be addressed in the operation. So the fuzzy relationship value of this rule can be represented as in Figure 4b.

**Rule 3:** Bending angle. This rule is to determine the fuzzy relationship value according to the angle of each plane. This is the angle between the mother plane and each rotated plane. If this angle is greater than 90°, the bending process is divided into one or more processes. The fuzzy relationship value is unity in the case of a bend angle less than 90° and zero in other cases. These relationships according to bend angles are represented as fuzzy functions as shown in Figure 4c.

**Rule 4:** Feeding direction. This rule is to determine the fuzzy relationship value of a fuzzy function according to whether or not the bend is in the feeding direction. After bending, an escape space is necessary in either the stripper plate of the upper die or the die plate of the lower die. The escape space should be at a minimum considering the die strength, the part to be fixed, the loss of die material, and the manufacturing time. Bending processes requiring a large escape space should be performed later to minimize the escape space. Because a bend perpendicular to the feeding direction requires a smaller escape space than a bend in the feeding direction, the former precedes the latter. The membership value for the perpendicular feeding direction is unity, and zero otherwise. The fuzzy membership function for this rule is shown in Figure 4d.

Note: To determine the grade of the membership for each group, the maximum grade of its bends is chosen and later used in the fuzzy matrix.

![Distance Rule](a.png)  ![Number of bends in a group](b.png)
Let $C=\{c_i\mid i=1, 2, \ldots, n\}$ represent the set consisting of all the remaining piercing operations or bend classes that are being considered for bending, where $c_i$ is one of the bend classes or piercing operations.

Let $R=\{r_j\mid j=1, 2, 3, 4\}$ represent the set of four criteria in the handling rules for bending or three rules in piercing operations sequencing, where $r_j$ represents one of the criteria.

A fuzzy relation is a mapping from $C \times R$ into $[0, 1]$, such that $(V_{ij})$, is expressed as follows:

$$(V_{ij}) = f(c_i, r_j)$$

Since related sets $C$ and $R$ are finite, a fuzzy relation $f$ on $C \times R$ can be represented as a fuzzy matrix $[M]$, the entries of which are $(V_{ij})$.

The determination of the grade, which may vary anywhere between zero and unity, is based on the sequencing rules. The fuzzy matrix $[M]$ for bending classes is shown in Table 1.

<table>
<thead>
<tr>
<th>Rule (r1)</th>
<th>Rule (r2)</th>
<th>Rule (r3)</th>
<th>Rule (r4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{11}$</td>
<td>$V_{12}$</td>
<td>$V_{13}$</td>
<td>$V_{14}$</td>
</tr>
<tr>
<td>$V_{21}$</td>
<td>$V_{22}$</td>
<td>$V_{23}$</td>
<td>$V_{24}$</td>
</tr>
<tr>
<td>$V_{31}$</td>
<td>$V_{32}$</td>
<td>$V_{33}$</td>
<td>$V_{34}$</td>
</tr>
<tr>
<td>$V_{41}$</td>
<td>$V_{42}$</td>
<td>$V_{43}$</td>
<td>$V_{44}$</td>
</tr>
</tbody>
</table>

### 7.2 Determination of Final Value Matrix (FVM) set

FVM is a matrix consisting of fuzzy values for each bend groups or piercing operations that are being considered. The fuzzy values are determined by implementing rules described in sections 5 and 6.2. These rules have been found to give suitable results in piercing and bending operations. The higher value an operation has, the sooner it should be performed. The relative importance of the rules is represented as a fuzzy set $W[R]$, as shown in equation 2 for the piercing operations and equation 3 for the bend groups.

$$W_{p}[R] = \{\eta_1 * 2.0, \eta_2 * 1.5, \eta_3 * 1.0\}$$

$$W_{b}[R] = \{\eta_1 * 1.2, \eta_2 * 0.8, \eta_3 * 0.6, \eta_4 * 0.2\}$$

Thus, the FVM set can be expressed as in equation 4:

$$FVM(C) = [V], W_{c/b}[R]$$

### 8. Die Layout Design

After the determining of operations sequence, the system determines number of the die stations and specifies which operations are performed in each station. According to these results, the system presents die stations layout as 2-D and 3-D model. User can interactively change operations sequence and add idle stations to suggested layout, if the proposed sequence is not suitable. At the end, the system presents the final layout. In part II of the paper, the system capabilities are explained using two sheet metal components.
9. EXAMPLE

Figure 5 shows a part used in automotive industry (Iran Khodro Company). This part has eight bends with several internal and external stamping features. When a component requiring piercing and intricate bending operations is applied to the system, the study considers the results carried out in each module.

9.1. INPUT AND SHAPE DEFINITION MODULE

The system automatically recognizes all of features such as bends, holes and notches. In this module first, material type and the thickness of component is extracted. Then bends and piercing features are recognized. Figure 6 shows bends parameters for this part extracted by the system. These parameters include radius, angle and length of each bend and their connecting planes. Other parameters of the bend feature such as bending direction, coordination of the start and end points are extracted too. According to bends lines vectors, the system suggests component arrangement on the strip.

Metal that is rolled shows grains along the direction in which the stock was drawn through the mill rolls. Bends should be at right angles to the grains or as close as possible to avoid breakage of the blank. So bends in each direction are calculated and the system suggests suitable feeding direction according to material type and number of bends in each direction. User can modified the components arrangement on the strip and change that. Figure 7 shows schematic of the part arrangement on the strip suggested by the system for this component and feeding direction according to arrangement. Optional arrangement which can be selected by user is shown too.

9.2. DESIGN FOR MANUFACTURING MODULE

The advisory system uses features data extracted in the recognition module. In the following some of the advisory system outputs are described. Figure 8 shows the slot features are not correct and the circle radius is
small. Also, the bends radiuses are small. These features are highlighted by red color and should be changed by the user. For example in the mother plane of the part, D-shape piercing feature is correct. The distances between features are enough and more than 2 times the thickness. But the diameter of the hole is less than 2.5 mm and the length to width ratio of the slots are more than 5 to 1 ratio. But in this example the ratio is 6.5. Also, according to material type (Hardened Low carbon steel) minimum radius of the bends is 6 mm. In the first step, the system highlights incorrect features and then the advisory module's comments are presented to the user.

Figure 8. Features checked by DFM rules

9.3 STRIP LAYOUT MODULE

In this module first, the mother plane is determined (it is highlighted in figure 5) and the flat view of part is created (Figure 9). Then in the sequencing stage, piercing and bending operations sequence are determined. This part has six internal piercing operations and eight external piercing (notches) operations. To separate the part from the strip, four cutting operations are used. Separating operations are defined interactively and ignored in the system.

Figure 9. Piercing and bending operations

According to sequencing procedure, first piercing operations are sequenced using fuzzy rules. Table 2 shows the results of fuzzy matrix and FVM for piercing operations.

Table 2. Fuzzy matrix [Mc] and FVM for piercing operations

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>FVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1</td>
<td>0.65</td>
<td>0</td>
<td>2.975</td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>3.9</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
<td>0.1</td>
<td>1</td>
<td>3.15</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
<td>0.92</td>
<td>1</td>
<td>4.38</td>
</tr>
<tr>
<td>C5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>C6</td>
<td>1</td>
<td>0.65</td>
<td>0</td>
<td>2.975</td>
</tr>
<tr>
<td>N1</td>
<td>0</td>
<td>0.62</td>
<td>1</td>
<td>1.93</td>
</tr>
<tr>
<td>N2</td>
<td>0</td>
<td>0.88</td>
<td>1</td>
<td>2.32</td>
</tr>
<tr>
<td>N3</td>
<td>0</td>
<td>0.75</td>
<td>1</td>
<td>2.125</td>
</tr>
<tr>
<td>N4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>N5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>N6</td>
<td>0</td>
<td>0.75</td>
<td>1</td>
<td>2.125</td>
</tr>
<tr>
<td>N7</td>
<td>0</td>
<td>0.88</td>
<td>1</td>
<td>2.32</td>
</tr>
</tbody>
</table>
According to table 2, number of the piercing stations is 9. Piercing operation C4 is punched in the first station and piercing operations N4 and N5 are performed in the last station.

Bending operations are sequenced in two stages first, they are classified and then the bending classes are sequenced by fuzzy logic.

According to the classification rules, the classes of this part are determined as follows:
- Bends B1 and B5 are in class 1, according to rule 2.
- Bends B2 and B6 are in class 2, according to rule 2.
- Bends B3 and B7 are in class 3, according to rule 2.
- Bends B4 and B8 are in class 4, according to rule 1.

After the classification of the bending operations is determined, the sequence of the operations can now be determined according to the fuzzy matrix as shown in table 3.

<table>
<thead>
<tr>
<th>Class</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>FVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>Class 2</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>Class 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.8</td>
</tr>
<tr>
<td>Class 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.4</td>
</tr>
</tbody>
</table>

According to table 3, number of the bending stations is 4. Bends B3 and B7 are performed in the first station and bending B1 and B5 are performed in the fourth station of the bending stations. Figure 10 shows the suggested die layout for this part. In this figure, some of the piercing stations are not shown, because if all stations are shown, figure quality would be reduced.

If this layout is compared with the die layout used in industry [12], it can be seen that the proposed bending operations sequence is similar to industrial die, but the number of proposed piercing stations is more than that. Thus piercing operations sequence should be changed by user. The user can change and modify the system output in system edit menu. He can add operational or idle station to die layout.

According to the number of piercing stations, in next study authors try to develop and improve the sequencing method for the piercing operations.
10. SUMMARY

In this paper a new system for the determination of bending and piercing sequences for progressive dies is described. The system includes three modules as follow: Input and shape definition, design for manufacturing, stations layout. In this system, sheet metal components are modeled using solidworks and their features data are extracted. Then features are controlled by DFM rules. In next step, operations are sequenced and 2-D and 3-D layout of the die stations are presented. In this system, first piercing operations and then bending operations are determined. Fuzzy set theory are used to determine piercing operations sequence, but determination of bending operations sequence are done in two stage; First the bends are classified then the bends classes are sequenced using fuzzy set theory. After sequencing, the system presents the part in each bending stations as 3-D model and die stations layout as 2-D and 3-D. The suggested layout can be modified interactively if it is not suitable.

REFERENCES

[12] The authors experience and observations in “progressive dies for Iran Khodro Automotive Company.” Tehran, Iran.
Implementation of a Modular Supervision System for a Flexible Manufacturing Cell

João C.E. Ferreira¹*, Hugo G. Santos², Marcelo T. dos Santos³, Léo Schirmer³

¹Univ. Federal de Santa Catarina
Depart. de Engenharia Mecânica
GRIMA/GRUCON, Caixa Postal 476, Florianópolis, SC 88040-900, Brazil

²Petrobras
Rua Felipe Musse, 803
Bairro Ubatuba
São Francisco do Sul, SC 89240-000, Brazil

³Brasilmatics Automação e Tecnologia Ltda.
Rua Arno Waldemar Döhler, 957
Zona Industrial Norte, Joinville, SC 89219-510, Brazil

ABSTRACT

This work proposes and implements a group of procedures to integrate physically and logically a group of CNC machines, an industrial robot and an automated storage and retrieval system (AS/RS) into a Flexible Manufacturing Cell (FMC). Initially, the modeling of the FMC was carried out, based on the definition of a group of minimum interfaces for each piece of equipment, through the use of Petri Nets. After modeling the cell, an integration model was implemented, based on management modules developed using a supervision and control software (Elipse SCADA). Each module was conceived with enough interfaces to allow the integration of different manufacturers' equipments, which provides flexibility to the configuration of the cell. The communication among the several modules is accomplished using the OPC (OLE for Process Control) technology, which enables the data exchange among managers in real time and in an open way. The integration between the managers and their respective pieces of equipment is made through PLCs (Programmable Logic Controllers). The system allows the manufacturing simulation of a family of parts, without the interference of machine operators. Human-Machine Interfaces (HMI) were developed to allow local and remote users (i.e. physically distant) to input and monitor their production orders in the cell.

1. INTRODUCTION

Manufacturing has been affected by profound changes in the last decades. The globalized competition, together with the increasing demand for product innovations, personalized products, high quality, and lower prices, force the companies to modernize their production methods. Changes used to be implemented as a means to provide a distinguishing characteristic to a company in relation to another, but now they are made as a survival factor in an uncompromising and competitive market. The distinguished technological characteristic that some corporations had in the past, which allowed the production of certain products with unique features, is becoming increasingly rare today. In the past, the consumer had to adapt to products supplied by the companies, but today the companies need to understand and adapt to the consumers' demands.

Manufacturing systems that prioritize production volume, with low cost and small variety, are facing the challenge to increase the flexibility of production and to diversify their products, and at the same time to decrease costs. For small and medium production batches, Flexible Manufacturing Cells (FMC) and Flexible Manufacturing Systems (FMS) are viable alternatives to respond to the challenges of flexibility, variety and reduced costs.

The concept of FMS involves a high degree of automation, where manufacturing is accomplished through automated machining centres, and the transfer among the machines is done by automated conveyor belts or automated guided vehicles. Positioning and motion of the parts are accomplished by robots. The raw material and the finished products are stored at an automated storage and retrieval system. The resources are linked to a computer system that coordinates all the actions, enabling the system to produce different types of parts. The technical complexity and the high acquisition costs and implantation of FMSs hinder its use significantly in companies and education and research institutions.

There are nowadays two possible solutions for the installation of a FMC or a FMS. The first of them is to acquire a turn-key solution supplied by companies such as Yamazaki Mazak [1], Kearney & Trecker [2], Heller [3], among others. In this case, technology transfer to the user does not usually happen, because the systems are

* Corresponding author: Tel.: +(55)(48) 3721-9387 ext 212; Fax: +(55)(48) 3721-7615; E-mail: jcarlos@emc.ufsc.br
not open. The second option is to build a FMC or FMS with the physical and logical integration of a group of pieces of equipment for processing, handling, transporting, and storage. In the latter case it is necessary to define a group of interfaces for the machines, to model the manufacturing system, to establish an integration pattern for the heterogeneous equipments, a communication pattern, a control system and a computer system that manages the whole structure and provides means to interface with the user. This work describes the integration of machines in a FMC in the latter approach, by applying a modular, open and expandable philosophy, in such a way that the FMC can manufacture different parts in an autonomous way.

2. EQUIPMENT IN THE FMC AND INTEGRATION ARCHITECTURE

The equipment used for the development of this work is located at Sociedade Educacional de Santa Catarina (SOCIESC) in Joinville, Brazil. It is composed of one CNC machining centre, one CNC lathe, an industrial robot, and an automated storage and retrieval system (AS/RS). Figure 1 shows a partial view of the machines being integrated. The computer resources are composed by two PLCs and six computers, all interconnected through an Ethernet network.

![Figure 1: Partial view of the machines being integrated in the FMC](image)

One of the most difficult problems to form a FMC consists of integrating different pieces of equipment, manufactured by different suppliers. Individually these machines present few communication resources with the outside environment. The robot, the CNC machines and the AS/RS have different hardware and no communication pattern. This work proposes the implementation of an open system, in which the different manufacturing resources are represented by modules, with a minimum group of interfaces. Managers are built to command the actions of the modules, in order to exchange data amongst themselves through an open standard for industrial communication, and that does not require the use of non-standard protocols.

The supervision and control software Elipse SCADA [4] will be used to build the managers, one for each resource in the cell. PLCs will make the connection among the managers and CNCs, which have only one serial port (RS232) for receiving the machining programs (G codes). The robot, which is equipped with a Profibus standard communication card, uses this resource to communicate with a PLC that is connected to the manager. The AS/RS has a human-machine interface (HMI) that is modified to accomplish the manager role. Figure 2 presents the proposed architecture for integrating the cell.

Besides the managers that correspond to the machines present in the cell, two other managers will be developed, with the objective of providing an interface with the users. One of them (FMC manager) should allow the local user to input his/her manufacturing orders, and to provide the user with all the information about the status of the machines, status of the manufacturing orders, to print reports, and to inform about any alarms. The second manager (remote manager) has the function of allowing an user who is physically distant from the cell to input his/her manufacturing orders.

3. INTERFACES

To accomplish the integration of the FMC, it is necessary initially that the different machines exchange data amongst themselves. The strategy used for interconnecting the pieces of equipment was the construction of one manager for each of them, using the Elipse SCADA supervision software. The different managers exchange messages using an Ethernet network. Each manager exchanges information with its respective equipment through PLCs.
To begin modelling the cell it is necessary to define the possible events, the hierarchy of the control system, and the minimum number of necessary messages to synchronize the actions in FMC. The definition of the necessary messages for modelling the cell is based on the combination of two main topics. Firstly, the choice of the control architecture that seeks to decentralize the decision process, increasing the autonomy of the system. All the managers built for integrating the cell act on the same level and in a cooperative way. The second point refers to the use of modular control through Discrete Events Systems (DESs), which allows complex problems to be decomposed into simpler modules, and then later the solutions are set up in a modular structure.

Torrico and Cury [5] proposed a model based on a modular hierarchical supervision control through state aggregation, where the hierarchy is divided into two levels: a level associated with the operator, and another level associated with manager. The application of this architecture allows a consistent and non-blocking behavior between the hierarchy levels, and the combined action of the several managers in the lower level. This model is applied in this work, where the level associated with the manager is represented by the equipment manager and the set of conditions and rules that compose the decision process. The second level is associated with the operator, and is composed by the execution elements, which are responsible for the effective integration between the managers and the pieces of equipment, accomplished through PLCs.

3.1. INTERFACES OF THE CNC MACHINES

The robot is responsible for loading and unloading the parts in the CNC lathe, CNC machining centre, and the CNC R1 (hypothetical third CNC machine). The cell manager is responsible for informing each CNC machine about which NC program (G code) should be loaded. There is no exchange of information between the AS/RS and the CNC machines.

After decomposing all of the possible actions of the CNC machines, and considering that each module manager should have the greatest possible autonomy, the interfaces were determined.

The CNCs should receive the information regarding which NC programs should be executed from the cell manager. The information on the robot states are supplied directly by the robot manager, and they indicate whether the robot is free or in some loading or unloading procedure, or if it placed or removed some part from the fixturing device of the CNC machines, and if the robot is within an area considered safe for the CNC machines to operate, avoiding the risk of collision. The information that the CNCs should supply the cell manager and the robot manager are basically to indicate whether the CNC machine is free or executing a machining process, if the door is open or closed, and if the fixturing device (three-jaw chuck, vice, collet, etc.) is open or closed.

With these defined and standardized interfaces, different CNC machines can be included in the FMC, provided they are physically within the robot workspace.

3.2. INTERFACES OF THE ROBOT

Still applying the philosophy of decomposing the robot actions in the FMC, and considering that the robot moves the parts between the AS/RS and the CNC machines, the group of interfaces can be defined. All the CNC machines (i.e. lathe, machining centre, and CNC R1) should provide the robot with indications whether they are idle or machining, the condition of their doors, and their fixtures. The AS/RS informs the robot about the presence of parts to be machined in the “mesa1” (table1) or in the “mesa2” (table2) of the machining cradles. The FMC manager should indicate which program the robot should execute.
The robot needs to inform its condition (idle or in loading movement), the position for loading or unloading the fixturing devices of the CNC machines, the retrieval or storage of parts in the pallet on mesa1 or mesa2 of the machining cradles, the presence within the area considered as safe for operation, and the condition of operational abnormality or emergency.

With these interfaces, the robot is capable of supplying and receiving the necessary information to load the appropriate programs and to synchronize all the events of material manipulation in the FMC.

3.3. INTERFACES OF THE AUTOMATED STORAGE AND RETRIEVAL SYSTEM

The AS/RS is initially responsible for supplying the raw material to be processed. After the machining, the AS/RS can store the finished product or to move it to the exit conveyor, for subsequent transport and delivery to the final customer. The AS/RS receives information from the FMC and robot managers. The FMC manager provides the information regarding which parts were selected for machining, the amount and the destination of the products already processed, as well as the strategy of movement of the transelevator. The return of the parts, coming from the CNC machines, to the mesa1 or mesa2, is signalled by the robot.

The information that the AS/RS needs to make available to the system is: (a) the pallet arrival with parts to be processed in the mesa1 or mesa2 of the machining cradles, and (b) the messages about the internal states of the AS/RS and the transelevator.

Besides the interfaces described above, there are others associated with the AS/RS, such as the entry and the storage of raw material. These interfaces are not analyzed based on the possible actions of the AS/RS, because they are not essential for the synchronism and integration of the cell. Those interfaces are responsible for the exchange of data between the AS/RS manager and its database, i.e., they are information treated in an autonomous way by the AS/RS manager, reducing significantly the traffic of messages among the different managers in the cell.

3.4. INTERFACES OF THE FMC MANAGER

The FMC manager has the primary function of serving as interface with the local and remote operator, allowing them to: input manufacturing orders, indicate the movement strategy, priority, amount, etc. Also, the manager should monitor the main events associated with the equipments in the cell, in order to determine and to communicate the choice of the NC programs and the robot programs. Hierarchically the FMC manager is on the same level as the other managers in the cell.

4. DISCRETE EVENT MODELING THROUGH PETRI NETS

In order to model a FMC, different tools can be used, among them the Finite Automata and Petri Nets [6]. According to Mušic and Matko [7], one of the greatest problems in the use of automata for developing the structure of supervision and control systems in real industrial processes, is the explosion of states. Thus, Petri Nets (PN) are advantageous, especially with regard to the possibility of modelling the aspects related to the synchronization and the non-determinism of the execution, or parallelism, and for presenting an additional information structure that reduces the problem of the explosion of states [8].

In this work, PN were used to model the FMC. One of the reasons for this choice is that PN describe in details the characteristics of the necessary conditions to trigger each transition, and the execution of the necessary actions to operate the system. This description of conditions and actions are transposed in an almost literal way for the generation of the supervision and control codes in the Elipse SCADA software, facilitating the construction of the managers extremely.

Figure 3 presents the PN structure that models and interprets the movement of the raw material from mesa1 of the machining cradle to the CNC lathe. The places represent the states of the resources; the transitions model the sequence of actions (events) that modify the resource states; the arcs represent the flow of the process; the tokens indicate that the predicate associated with the place is true. The return of the part already processed by the lathe to the present pallet in mesa1 is described by other PN. That PN is composed of 12 places, 9 transitions, and 16 variables.

5. CONVERSION OF THE IPN TO SUPERVISION CONTROL CODES

From the cell model using IPN (Interpreted Petri Net), the trigger conditions of each transition and the actions resulting from triggering can be directly translated to programming scripts of the Elipse SCADA software. The variables used in the modelling that indicate the conditions and the actions associated with each transition are already in the format of tags.
Implementation of a Modular Supervision System for a Flexible Manufacturing Cell

Figure 3: Simplified IPN for moving parts from table1 to the CNC lathe

Ellipse SCADA is an event-oriented software, which means that the execution of any programming script depends directly on the occurrence of a discrete phenomenon. This characteristic is used as trigger condition of each transition in the cell model. The programming script, always associate with an event, contains the codes that represent the actions to be executed by the system. The control of the evolution of the states in the supervising system is centred on the events that trigger the transitions. Figure 4 presents an example of the conversion of the IPN model for generating the supervision controls, and they represent the update of the NC programs in the cell, triggered by the arrival of a pallet in mesa1 of the machining cradles of the AS/RS (OnValueChanged event - it triggers whenever its value changes).

6. MANAGERS OF THE STORAGE SYSTEM AND THE FMC

The implementation of the different managers proposed in this work was done similarly, and therefore this paper describes the implementation of the managers of the AS/RS and the FMC. It is intended to build these managers within a single application, unifying the input and output interfaces. Initially, the conversion to tags is carried out, which is the first step for constructing the application.

The AS/RS manager should command the following actions: entry of the pallet containing raw material or finished parts, storage of pallets, transportation of parts to the machining cradles, return of the parts from the machining cradles, and pallets retrieval. The intervention of an operator consists of loading the material on the pallets and their placement on the entry conveyor. This task can also be accomplished by an AGV, and the application of the AS/RS includes this condition, but it will not be used in the present implementation because the laboratory does not have an AGV. When accomplishing the entry of a raw material in the AS/RS, the operator needs to inform a code that refers to the material type to be stored. These codes are previously registered in the manager's database, as well as the machining code, which needs a corresponding raw material and a pre-determined machining table. The database used is Microsoft Access.

To order the production of a part, the user should input the code of the part, in order to indicate the amount of parts to be produced, where the part will be moved to after the return from usinagem, and the priority of the operation (normal or urgent).

The tag “MovPeca” (move part), belonging to the AS/RS manager, is associated with an object type Setpoint, where the user types in the code of the part to be machined. When the user presses the Confirm key, the content of the MovPeca tag is copied to several tables of the database. When the task is executed, and the pallet arrives at one of the tables of the machining cradles, the tag “ProgUsinagem” (machining program) is attributed the value of the MovPeca tag, through a reading of the database. The box “Saida” (exit) indicates that the pallet will be moved to the exit cradle, and not to the storage shelves, after the machining process is completed. Figure 5 shows the script associated with the tag “presenca_mesa1” (present on table 1), triggered by the arrival of the pallet in mesa1, associated with the OnValueChanged event. This script updates the ProgUsinagem tag that will be read by all the pieces of equipment in the cell through OPC (OLE for Process Control) communication standard. Depending on the reading of this tag, the robot manager will load the movement program “m1f1” (from table 1 to machining centre 1), and the machining centre will also load the corresponding machining program.
Figure 4: Example of conversion of IPN into control code

Figure 5: Script associated with the presence_table1 tag
The instruction of the script "ProgUsinagem=tab05_usinagem.produto" is responsible for updating the ProgUsinagem tag from "tab05_usinagem", "product" column in the database. The machining table in the database is used by the AS/RS manager to rank machining tasks. The user can input several machining orders at the same time, and each order will be inserted in the list of tasks in the database. The process of updating the ProgUsinagem tag, from the machining table, enables the synchronization of the robot movement and the load of the program in the CNCs with the arrival of the pallet containing the raw material in the machining tables. The tag “código_retorno_M1” (code for returning M1) receives a numeric value that indicates the CNC machine from which the part will return to mesa1, and the phase of the process in which the part is. The first digit indicates the CNC that is processing the order (lathe=1, machining centre=2, and CNC R1=3). If the second digit is equal to zero, it means that the part is being load in the CNC machine.

After updating the ProgUsinagem tag, caused by the arrival of the pallet in mesa1, the robot executes the program m1f1, and moves the raw material to the machining centre. When the machining of the part begins, the ProgUsinagem tag should be updated immediately, otherwise the program containing the motion of the robot would repeat the initial request. To accomplish this function, the closing of the door of the machining centre (which starts the machining cycle) is used to update the ProgUsinagem tag. In this way, when the robot returns for “Phome”, it will be capable of executing a new requested movement.

The AS/RS manager, as well as the FMC manager, needs many pieces of information about the other managers, as defined by their interfaces. Figure 6 shows the structure of the OPC servers in the organizer of the application and in one of the screens of the FMC manager.

Figure 6: Structures of the OPC servers and a screen capture of the FMC manager

The Manager screen provides the user with a summary of the states of the resources in the cell. The information about the CNC machines, the robot, and of the remote manager is obtained from OPC servers. On the other hand, the information about the AS/RS is directly linked to the tags of the AS/RS manager. The OPC tags are decomposed through scripts and organized in a group of tags ("Pulgus"), which contains the following groups of secondary tags: Auxiliaries, CNC R1, Machining Centre, Robot, and Lathe.

7. GENERAL INTEGRATION OF THE FMC

In this combination of the managers of the manufacturing resources with the FMC manager and the remote manager, all of the managers are connected through a local Ethernet network, and each manager carries out the OPC server and client roles at the same time. When an application supplies information to other managers, it is an OPC server, but when it receives data from another manager, it behaves as an OPC client.

Several versions of the Windows operating system were used in the managers (e.g. NT, 2000, XP-HOME and XP-SP2), which executed in several computers, with Pentium 200 MMX to Pentium 4 2.8 MHz processors. All the computers have Ethernet cards, and those that communicate with PLCs have at least a RS232 serial port, except the FMC/AS-RS manager, which has two serial ports (one used to communicate with the master PLC Moeller, and other with the bar code scanners of the pallets). The robot controller has a Profibus network card (DSQC 352 model), used to communicate with the Siemens PLC.
After integrating the managers, the FMC is capable of producing the parts previously registered in the database of the FMC/AS-RS manager. To begin the operation of the system, it is necessary that all the managers are in operation, the communication network is active, and all the pieces of equipment and PLCs have their programs loaded.

The user interface with the FMC is the FMC/AS-RS manager. In this HMI the user registers the parts and processes, inputs the entry and exit of raw material in the AS/RS, and requests the production of the parts previously registered in the system. The access of the user to the manager is accomplished through password, for safety reasons. The FMC/AS-RS manager also allows monitoring the main events and states of the manufacturing resources in the cell, to emit reports on produced and stored parts, and to find stored raw material and finished parts.

The ABB robot, model IRB 2400/10, has a claw for manipulating and transporting materials between the AS/RS and the CNC machines, and vice-versa. Any other robot type can be used in the cell, as long as its workspace reaches the tables of the cradles of the AS/RS and the CNC machines. Also, it should have an appropriate claw for manipulating and transporting materials in the cell, and its controller should have sufficient capacity for programming and communication with the external environment.

The used CNC machines, a Feeler FTC10 lathe and a Feeler FV-600 machining centre, demanded hardware and software adaptations for proper integration in a FMC. The necessary alterations encompass the automation of the doors and fixtures, besides the reservation of a group of internal memories to indicate the download of the NC machining programs. Other CNC makes and models can be used in FMC.

To include a new CNC machine in the cell, it is not necessary to alter the model of the cell, being enough to build a new manager similar to the lathe, machining centre, and R1 managers. Also, the programs for moving the robot should be generated for tending the new CNC machine, and the programming scripts associated with the OPC tags should be altered.

In the case of including a measurement equipment, which could be a laser micrometer, it would be necessary to define a group of interfaces for building the new manager. In this case, once again it would not be necessary to model the cell. The robot movement would have to be altered to take the part to the measurement system. If the part is approved, the robot would move the part back to the pallet. In case of rejection, the robot would move the part to a reject or rework buffer.

8. CONCLUSIONS

This work aimed at providing a guideline for constructing a FMC through the physical and logical integration of heterogeneous pieces of equipment, and we consider that this objective was reached. The proposal encompasses the modelling of the system using Interpreted Petri-Nets, the specification of a set of interfaces for each equipment, the construction of managers based on a SCADA supervision software, the integration of the managers with CNC machines and robot using PLCs, and the integration of the managers through the OPC industrial communication standard.

As future works, it is necessary to develop a new FMC manager, using the Elipse E3 supervision software, which enables to incorporate all the managers developed with Elipse SCADA, besides making it possible the remote operation of the FMC through the Internet.

REFERENCES

Cooperative robot kinematics to perform load sharing tasks

Cristiane P. Tonetto¹ and Altamir Dias

Mechanical Engineering Department
CAD/CAM Laboratory
Santa Catarina Federal University
Florianópolis, Santa Catarina, 88040-900, Brazil

ABSTRACT

This paper presents a research of robot's cooperation applied to load sharing tasks, by using non usual techniques such as Kirchhoff-Davies method, Assur virtual chains and screw theory. Many robotics applications are currently being restricted by the fact that there are limits in the performance of tasks by a single robot. One example of such restriction is when it is necessary to move one object from a point to another and it is very heavy or extremely big to be done by only one robot. In this case, two or three robots are needed in order to achieve the task. So, the paper focus will be centered in the description of the task trajectory on the operational space, and how to synchronize and establish the cooperation between two or more robots, by using non usual techniques. It explores some procedures to describe how the robots can move over the trajectories needed complying with the task(s) parameters. By using the screw theory it is possible to know and to control the relative movements between all the manipulators links. A procedure based on the Kirchhoff-Davies method is used to establish the instant speeds between joints from the desired speeds/positions on the robots workspace, thus being possible to solve the inverse kinematics of the robots. The kinematic chain of the robots is considered as closed chains, so additional auxiliary virtual chains are added to the system in order to close opened chains. Virtual chains, also known as Assur virtual chains, are used to define desired characteristics or constraints on the movements to be performed by the robots, besides being used to couple the robots movement. All the trajectory planning is intended to be planned off-line. Although this approach is new and uncommon on the current literature, the final results are very good and compatible with the results of other techniques. Finally, a simulation sample of robots performing tasks on the operational space is presented.

1. INTRODUCTION

Due to the ever growing request for industrial products and thus higher production, robots are more and more commonly used in the industry. Eventually the need to use many robots cooperatively emerges, such as in part handling and load sharing operations. A cooperative robots system is defined as a system composed by several robot manipulators, acting collectively in order to perform one or more tasks. So, to program cooperative robots it is necessary to know how to generate automatically the robots program to different configurations, what is its geometry and how to give the task specifications. Some authors have researched the use of more than one robot performing tasks, such as Lewis [1] and Owen, Croft and Benhabib [2], which assume the whole system as a single redundant robot and the way to plan trajectory for two robots; Dourado [3] researches the kinematic and trajectory planning for planar cooperative robots and Ribeiro, Guenther and Martins [4] propose a Cooperation Jacobian for cooperative robots.

This paper intends to present some basic tools used on the robots kinematics computation and further describes the procedure to compute kinematics of two cooperating robots performing tasks such as the load sharing tasks.

2. TOOLS FOR ROBOT’S KINEMATICS COMPUTATION

The differential kinematics relates the joint speeds to the end effector’s, and can be represented, for example, as a Jacobian matrix (J) that is a function of the current configuration of the robotic manipulator. The Jacobian for serial

¹ Corresponding author: Tel.: +55(48) 3234-7186; E-mail: cris.tonetto@emc.ufsc.br
Flexible Automation and Intelligent Manufacturing, FAIM2009, Teesside, UK

manipulators is usually attained by using one of two methods: one based on the Denavit-Hartenberg convention and other on the screw theory [5].

This paper uses the screw theory approach, as it is a new methodology, successfully used on several robotics researches, such as on the UFSC robotics group. On the other side, the Denavit-Hartenberg is well known and yet widely used on robotics researches.

2.1. SCREW THEORY

The Chasles’ theorem establishes that the displacement of a rigid body can be represented as a rotation around a fixed axis and a translation in the direction of the same axis. The combination of the translation and rotation movements is called screw displacement [6]. As the screw displacement can represent both rotation and prismatic displacements, it is possible to represent two different types of joints that define the relative displacements of links in a single way, so being possible to overcome pose questions in a generalized way.

Figure 1 depicts the screw displacement of a point $P$ on the space from the $P_1$ to the position $P_2$. These displacements can be represented by a rotation of an angle of $\theta$ over the screw axis followed by a translation of $t$ on the same axis. The vector $p_i$ denotes the initial position of the point $P$ referenced to the coordinate’s system $O$, the vectors $r_1$ and $r_2$ are same size vectors that define the initial and final position of $P$ according to the screw displacement. The $p_f$ denotes the position of $P$ after the translation, according to the reference system $O$ [6].

![Figure 1: Displacement of the point $P$ on space](image)

On a system where many rigid bodies are displaced, the $s_i$ vector denotes the unitary vector that defines the direction of rotation and translation of the screw axis, so that $i$ refers to the index of the link or fixed body. The $s_{oi}$ vector denotes the $s_i$ vector of the position according to a fixed coordinate system of a reference link. The angle $\theta$ is the rotation angle of the $P$ point and $t$ is the length of the displacement of the translation of $P$ [6,7]. For convenience, the vectors $s_i$ and $s_{oi}$ shall be chosen perpendicular, that is:

$$s_i^T \cdot s = 0 \quad (1)$$

The Mozzi theorem states that the speed of the points of a rigid body relative to a reference system $O(x,y,z)$ can be represented by a differential rotation around a fixed axis and a simultaneous differential translation on the same axis [5]. Thus, any movement of a rigid body can be considered a screw displacement.

To take in account that the instantaneous movement of a rigid body relative to a inertial system is composed by a pair of vectors $(\omega, v_p)^T$. So, $\omega=(\omega_x, \omega_y, \omega_z)$ denotes the angular speed of a body relative to the chosen coordinate system and the vector $v_p=(v_{px}, v_{py}, v_{pz})$ represents the linear speed of a point $P$ that is in the body [5].

Once more, the instantaneous movement can be decomposed as an amplitude and a normalized axis, when there is only rotational or translational movement. When the move is a rotation, the amplitude is the magnitude of the angular velocity $||\omega||$, and when the move is a translation, the amplitude is the magnitude of the linear velocity $||v_p||$. 

Page 131
A screw $\hat{s}$ is the geometric element defined by a directed line on the space and by a scalar parameter $h$ of length unit that defines the pitch. The normalized screw $\hat{s}$ of a joint $i$ is given by:

$$\hat{s} = \begin{bmatrix} s_i \\ s_o \times s_i + h \cdot s_i \end{bmatrix}$$  \hspace{1cm} (2)$$

where $h$ is the screw pitch and $s_o \times s_i$ is the vector product between $s_o$ and $s_i$ vectors.

It is possible to verify that when the rotation movement is around the $s_i$ axis and the translation over the $s_i$ axis, the screw represents the instantaneous movement:

$$\hat{s} = (\omega, v_p) = \hat{s} \cdot \dot{q}$$  \hspace{1cm} (3)$$

The screw can define the differential displacement between two bodies according to a reference coordinate system. The differential kinematics between two links of a given kinematic chain is generally defined by one revolute or translational joint. In the case of different types of joints, such as spherical joints, they can be decomposed in three different revolute joints.

When the movement is related to a revolute joint, there is no translation, that is, the pitch of the screw is null ($h=0$), and the screw displacement is given by the equation:

$$\hat{s} = \begin{bmatrix} s_i \\ s_o \times s_i \end{bmatrix}$$  \hspace{1cm} (4)$$

In the case of a prismatic joint, there is no rotation movement ($\theta=0$), only translative, so the pitch of the screw is infinite ($h=\infty$), and the screw displacement is given by:

$$\hat{s} = \begin{bmatrix} 0 \\ s_i \end{bmatrix}$$  \hspace{1cm} (5)$$

The resulting screw of the displacement of the end effector can be given by adding the screws of each one of the joints, that is:

$$\hat{s}_E = \sum_{i=1}^{n} \hat{s}_i \cdot \dot{q}_i$$  \hspace{1cm} (6)$$

where $\hat{s}_i$ is the $i^{th}$ normalized screw described in relation to the reference coordinate system, $\dot{q}_i$ is the magnitude of the displacement of the rigid body $i$ and $\hat{s}_E$ is the resulting screw that represents the end effector movement according to the reference coordinate system. Then the displacements of the joints represented by normalized screws and their magnitudes are computed [6].

### 2.2. Kirchhoff-Davies Method

The Kirchhoff mesh rule establishes that the algebraic sum of all potential differences on a closed electric circuit equals null. Davies adapted this mesh rule for mechanism kinematics computation on closed kinematic chains, so that the sum of relative speeds between two adjacent links throughout any kinematic chain is null. This is known as the Kirchhoff-Davies method, and it become possible to establish an instantaneous relationship between the speeds of all joints of a closed kinematic chain using the speeds representation instead of screws. This relation is named equation of manipulators kinematic chain constraint [3,5,7].

The application of the Kirchhoff-Davies method is better understood with an example with a $n$ bars and $n$ joints mechanism (Figure 2). So, suppose that the screw $\hat{s}_A$ represents the movement of link 2 related to the link 1, thus, the screw $\hat{s}_i$ represents the movement of the link $s_{i+1}$ related to the $i$ link, so that $1 < i < n$. The screws $\hat{s}_A, \hat{s}_B, \hat{s}_C ... \hat{s}_n$ represent the kinematics pairs $A, B, C, ... n$, respectively.
The link 2 motion related to link 1 is represented by the screw $S_A$, and the link 3 movement related to link 1 is represented by the algebraic sum of the screws $S_A + S_B$. As the link 1 is fixed to the ground its motion related to itself is null and can be written, as shown in the equation (7), applying the Kirchoff-Davies method for closed chains.

$$S_A + S_B + S_C + ... + S_n = \vec{0} \quad (7)$$

The $\vec{0}$ vector is a vector of (3×1) dimension. Changing its screws to its normalized axis $\hat{S}$ and by the speed magnitude $i_q$ leads to:

$$\hat{S}_A \hat{q}_A + \hat{S}_B \hat{q}_B + \hat{S}_C \hat{q}_C + \cdots + \hat{S}_n \hat{q}_n = \vec{0} \quad (8)$$

where $\hat{S}$ represents the normalized screw of the screw $S_i$ and $\hat{q}_i$ represents the magnitude of the speed related to the $i$ joint, so that $i \in \{A, B, C, ..., n\}$.

The matrix format is:

$$\begin{bmatrix} \hat{S}_A & \hat{S}_B & \hat{S}_C & \cdots & \hat{S}_n \end{bmatrix}_{3 \times 4} \begin{bmatrix} q_A & q_B & q_C & \cdots & q_n \end{bmatrix}_{4 \times 1} = \vec{0} \quad (9)$$

And generally, the constraint equation is:

$$N \dot{q} = \vec{0} \quad (10)$$

The equation (10) can be rewritten in order to find a way to solve the differential kinematics on the joint space of the kinematic chain and this way finding the speed magnitudes of the passive joints as a function of the actuated joints. The matrices of equation (10) can be divided in two different groups: the joints with known speed magnitudes (primary joints) and on the ones with unknown magnitudes (secondary joints), that is:

$$\dot{q} = \begin{bmatrix} \dot{q}_s & \vdots & \dot{q}_p \end{bmatrix}^T \quad (11)$$

Reassembling the $N$ matrix according to the magnitudes leads to:

$$N = \begin{bmatrix} N_s & \vdots & N_p \end{bmatrix} \quad (12)$$

Finally resulting on equation (13):

$$\begin{bmatrix} N_s & \vdots & N_p \end{bmatrix} \begin{bmatrix} \dot{q}_s & \vdots & \dot{q}_p \end{bmatrix}^T = \vec{0} \quad (13)$$

Setting the matrices apart produces equation (14):
Cooperative robot kinematics to perform load sharing tasks

\[ N_s \dot{q}_s = -N_p \dot{q}_p \tag{14} \]

And thus, equation (9) can be rewritten to:

\[ \begin{bmatrix} \hat{S}_{m+1} & \hat{S}_{m+2} & \cdots & \hat{S}_s \end{bmatrix} \begin{bmatrix} q_{m+1} & q_{m+2} & \cdots & q_s \end{bmatrix}^T = -\begin{bmatrix} \hat{S}_p & \hat{S}_q & \cdots & \hat{S}_n \end{bmatrix} \begin{bmatrix} \dot{q}_p & \dot{q}_q & \cdots & \dot{q}_n \end{bmatrix}^T \tag{15} \]

where \( m \) is the number of primary joints.

If \( N_s \) is invertible, the speed magnitudes of the secondary joints \( \dot{q}_s \) can be computed by the following equation:

\[ \dot{q}_s = -N_s^{-1} \cdot N_p \dot{q}_p \tag{16} \]

2.3. ASSUR VIRTUAL CHAINS

The Assur virtual chains were developed in order to help the application of Kirchoff-Davies method. An Assur virtual chain is a serial kinematic chain with virtual links and joints [5]. It must comply with certain properties, so that the virtual joints movements are represented by linear independent screws and a virtual chain cannot change the degree of freedom of a real kinematic chain.

Therefore, in order to use a kinematic chain, a serial chain must be defined, and it must have the same degree of freedom of the screw system in which the real kinematic chain is presented.

As the kinematic chains are modified by the added virtual kinematic chains without changing the original chains mobility, it is possible to get information and introduce characteristics on its movement, making possible the application of closed chains algorithms in open chains.

The virtual kinematic chains can be classified according to the screw system order \( \lambda \) in which the real kinematic chain is included. If the real kinematic chain is planar, the screw system has \( \lambda=3 \) and if the real kinematic chain is spatial the screw system has \( \lambda=6 \). The planar virtual chains can be PPR with two prismatic and one revolute joint or RPR with one prismatic and two revolute joints. The spatial virtual chains can be PPcP with three prismatic and one spherical joint, RPPs with one revolute, two prismatic and one spherical joint [5].

The definition of which virtual joints to choose depends over the desired characteristics of movement, defined by the space over which the trajectory will be defined. In spatial virtual kinematic chains, the order of the screw system is \( \lambda=6 \), and thus the virtual chain have a degree of freedom of six. The spherical joints can be swapped by three orthogonal revolute joints.

3. IMPLEMENTATION

In order to verify the described theory, two cooperative anthropomorphic robots were simulated, in a load sharing task, following circular trajectories. The size of the links, types of joints or robot model, the relative distance between robots and initial pose of the robots are considered inputs for the problem. From the robots model or links size it is possible to find all the values of \( s \) and \( s_o \) and to find the robot screws and to apply the Kirchoff-Davies method to compute the inverse kinematics.

It will be used two similar anthropomorphic robots, with spherical wrist (according to model in Figure 3). This robot has six revolute joints and its end effector has to follow the trajectory \( x=1,81+\cos(t), y=sin(t) \) and \( z=3,77 \) with \( 0 \leq t \leq 2\pi \). The initial configuration of each of the first robot joints are: \( \theta_1=0, \theta_2=7\pi/10, \theta_3=-\pi/2, \theta_4=-\pi/3, \theta_5=0 \) and \( \theta_6=0 \). The end effector of second robot must follow the movements of the first, so that the relative movement between the first and the second robot remains null. The initial configuration of the second robot is: \( \theta_7=\pi, \theta_8=7\pi/10, \theta_9=-\pi/2, \theta_10=-\pi/3, \theta_11=0 \) and \( \theta_{12}=0 \).
In order to find the values of $s_a$ and $s_o$, the reference position of the robot is considered (Figure 3b). As the planning is made off-line, the reference position can be any desired position, so it was chosen a position in which the robot was horizontally so that it is possible to visualize $s$ and $s_o$ vectors. The position of $s_o$ can be any point over the screw axis, and thus for convenience, it is chosen the point that lies over the robots joint. After that, the correct homogeneous transformations are applied over $s$ and $s_o$, in order to get the values of $s$ and $s_o$ of the manipulator in any configuration besides the reference position.

The values of $s$ and $s_o$ for the first stretched anthropomorphic manipulator are described on Table 1, the joints of the real manipulator are numbered from 1 to 6.

<table>
<thead>
<tr>
<th>Joints</th>
<th>$s$</th>
<th>$s_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,0,1)</td>
<td>(0,0,0)</td>
</tr>
<tr>
<td>2</td>
<td>(0,-1,0)</td>
<td>(0,0,0)</td>
</tr>
<tr>
<td>3</td>
<td>(0,-1,0)</td>
<td>($a_3$,0,0)</td>
</tr>
<tr>
<td>4</td>
<td>(0,-1,0)</td>
<td>($a_3+a_4$,0,0)</td>
</tr>
<tr>
<td>5</td>
<td>(0,0,1)</td>
<td>($a_3+a_4+a_5$,0,0)</td>
</tr>
<tr>
<td>6</td>
<td>(1,0,0)</td>
<td>($a_3+a_4+a_5$,0,0)</td>
</tr>
</tbody>
</table>

Table 2 describes the values for the joints of the virtual chain $P_x$, $P_y$, $P_z$, $R_x$, $R_y$ and $R_z$ that is added in order to close the kinematic chain started by the serial manipulator, so that it is possible to apply the Kirchhoff-Davies method.

<table>
<thead>
<tr>
<th>Joints</th>
<th>$s$</th>
<th>$s_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_x$</td>
<td>(1,0,0)</td>
<td>---------------</td>
</tr>
<tr>
<td>$P_y$</td>
<td>(0,1,0)</td>
<td>---------------</td>
</tr>
<tr>
<td>$P_z$</td>
<td>(0,0,1)</td>
<td>---------------</td>
</tr>
<tr>
<td>$R_x$</td>
<td>(1,0,0)</td>
<td>(0,0,0)</td>
</tr>
<tr>
<td>$R_y$</td>
<td>(0,1,0)</td>
<td>(0,0,0)</td>
</tr>
<tr>
<td>$R_z$</td>
<td>(0,0,1)</td>
<td>(0,0,0)</td>
</tr>
</tbody>
</table>

In order to compute the values of $s$ and $s_o$ from the data of the stretched manipulator an algorithm was implemented. This algorithm applies the homogeneous transformations over $s$ and $s_o$ for each joint recursively. Thus it is possible to have the values of $s$ and $s_o$ for each one of the special manipulators joints, in a generic way, as a function of the joints angles. On the implemented algorithm the resulting $s$ and $s_o$ vectors are a function of the angles $\theta$, that is, a function of the joint angles. The values of $s$ and $s_o$ that result from the application of the algorithm for the first manipulator, are given on the Tables 3 and 4.
Table 3: Values of \( s \) for the first manipulator.

<table>
<thead>
<tr>
<th>Joints</th>
<th>( s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,0,1)</td>
</tr>
<tr>
<td>2</td>
<td>(\sin \theta, -\cos \theta, 1)</td>
</tr>
<tr>
<td>3</td>
<td>(\sin \theta, -\cos \theta, 1)</td>
</tr>
<tr>
<td>4</td>
<td>(\sin \theta, -\cos \theta, 1)</td>
</tr>
<tr>
<td>5</td>
<td>((\cos \theta \sin \theta 1 - \cos \theta \cos \theta \cos \theta \sin \theta 1) \sin \theta 1 - (\cos \theta \sin \theta 1 + \cos \theta \sin \theta 1) \cos \theta 1)</td>
</tr>
<tr>
<td>6</td>
<td>(((\cos \theta \sin \theta 1 - \cos \theta \cos \theta \cos \theta \sin \theta 1) \sin \theta 1 + (\cos \theta \cos \theta \cos \theta \sin \theta 1) \cos \theta 1 + \cos \theta \sin \theta 1)</td>
</tr>
</tbody>
</table>

Table 4: Values of \( s \) for the first manipulator.

<table>
<thead>
<tr>
<th>Joints</th>
<th>( s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( s_{01} = (D_{11}, D_{12}, D_{13}) )</td>
</tr>
<tr>
<td>2</td>
<td>( s_{02} = (D_{21}, D_{22}, D_{23}) )</td>
</tr>
<tr>
<td>3</td>
<td>( s_{03} = (\cos \theta \cos \theta \sin \theta 1 + D_{13}, (\sin \theta \cos \theta) \alpha_2 + D_{13}, \sin \theta \alpha_2 + D_{13}) )</td>
</tr>
<tr>
<td>4</td>
<td>( s_{04} = (\cos \theta \cos \theta \cos \theta \sin \theta 1 + (\cos \theta \cos \theta) \alpha_2 + D_{13}, (\sin \theta \cos \theta) \alpha_2 + D_{13}) )</td>
</tr>
<tr>
<td>5</td>
<td>( s_{05} = (\cos \theta \cos \theta \cos \theta \sin \theta 1 + (\cos \theta \cos \theta) \alpha_2 + D_{13}, (\sin \theta \cos \theta) \alpha_2 + D_{13}) )</td>
</tr>
<tr>
<td>6</td>
<td>( s_{06} = s_{05} )</td>
</tr>
</tbody>
</table>

The Figure 4 shows a robot and the virtual chain used in the computation of the kinematics using the Kirchhoff-Davies method. It is considered that a 3P3R virtual chain links the manipulators base with the end effector, and the chain compounded by the end effector and the robot is closed and the Kirchhoff-Davies method can be applied.

Figure 4: Circuit that represents two collaborating manipulators
From the circuit it is possible to find the $N$ matrix:

$$
N = \begin{bmatrix}
    \dot{s}_1 & \dot{s}_2 & \dot{s}_3 & \dot{s}_4 & \dot{s}_5 & \dot{s}_6 & -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{R1} & -\dot{s}_{R1} & -\dot{s}_{R1} & \ldots \\
    0 & 0 & 0 & 0 & 0 & 0 & \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{R1} & \dot{s}_{R1} & \dot{s}_{R1} & \ldots \\
    \ldots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
    \ldots & \dot{s}_{p2} & \dot{s}_{p2} & \dot{s}_{R2} & \dot{s}_{R2} & \dot{s}_{R2} & -\dot{s}_7 & -\dot{s}_8 & -\dot{s}_9 & -\dot{s}_{p10} & -\dot{s}_{11} & -\dot{s}_{12} \\
\end{bmatrix}
$$

(17)

The $N$ matrix can be decomposed in $N_s$ and $N_p$:

$$
N_s = \begin{bmatrix}
    \dot{s}_1 & \dot{s}_2 & \dot{s}_3 & \dot{s}_4 & \dot{s}_5 & \dot{s}_6 & 0 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 0 & 0 & -\dot{s}_7 & -\dot{s}_8 & -\dot{s}_9 & -\dot{s}_{10} & -\dot{s}_{11} & -\dot{s}_{12} \\
\end{bmatrix}
$$

(18)

$$
N_p = \begin{bmatrix}
    -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{p1} & -\dot{s}_{R1} & -\dot{s}_{R1} & 0 & 0 & 0 & 0 & 0 & 0 \\
    \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{p1} & \dot{s}_{R1} & \dot{s}_{R1} & \dot{s}_{R1} & \dot{s}_{R1} & \dot{s}_{p2} & \dot{s}_{p2} & \dot{s}_{R2} & \dot{s}_{R2} \\
\end{bmatrix}
$$

(19)

The joints speeds are given by the equation:

$$
\dot{q}_s = -N_s^{-1}N_p \dot{q}_p
$$

(20)

Figure 5 depicts the simulated robots performing the task, with many joints position drawn. Only the manipulator on the left receives the trajectory and the right robot follows the movement, starting from another position over the piece, simulating a load sharing task, so that every robot grabs on a different place of the piece and movement the piece cooperatively. The implementation of the algorithm was executed in Octave [8]. The output of the software is the trajectory of the joints of both robots and the input, besides the robots data, is the desired trajectory defined as the primary joints $\dot{q}_p$ of the virtual chains:

$$
\dot{q}_p = \begin{bmatrix}
    -\sin(t) & \cos(t) & 0 & 0 & 0 & 0; & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}^T
$$

(21)

Figure 5: Two robots in cooperation

So, the trajectory programming is performed automatically, for the first robot by the reference path and for the second robot as null reference, defining that, according to the manipulated object, the end effector of the second robot must remain static. In order to comply with the circuit, the movement of the second robot is automatically computed, as a function of the movement of the first one, thus showing the need of cooperative work.
4. SUMMARY

The application of the described methods, as the screw theory, Kirchhoff-Davies, Assur virtual chains, allows the solution of the cooperation problem in a compact and simple way. The screw theory is able to model in a general way prismatic and revolute joints and operate over them, as illustrated by the spatial example for two anthropomorphic robots. The model of virtual chains is very flexible and allows the robot programmer to define the virtual chains that bests represent the input data. It is possible, for example, to use such concepts to model other kind of restrictions, such as the collision avoidance, as virtual chains. But it is object of research for future articles.

REFERENCES

On parametric toolpath design of a robot serving a press-brake

Nikolaos Kontolatis, George-Christopher Vosniakos*, and Kostas Kyriakopoulos

School of Mechanical Engineering
National Technical University of Athens
Zografou, Athens, GR-157 80, Greece

ABSTRACT

This work examines a robot cell consisting of a press brake and a robot which handles the sheet metal, thus replacing the human operator. The motion of both robot and press-brake ram is simulated in a CAD environment enhanced, first, with software simulating sheet bending kinematics and, second, with a genetic algorithm determining optimum bending sequence, both developed in previous research stages. The corresponding robot motion steps are configured following a task-based approach. They are defined parametrically in such a manner as to draw up relatively simple trajectories involving, typically, no more than three joints simultaneously. The range of applications involves variable part dimensions of the final part and of the initial sheet profile. The user is actively involved in several phases of the whole process, e.g. in determining the grasping point of the sheet, in requesting collision detection when required, in modifying robot path when necessary, etc. The methodology and tools developed constitute a good experimentation base towards the ultimate total replacement of the human by the robot.

1. INTRODUCTION

Automated handling in manufacturing processes may use hardware such as robotic arms and software such as CAD-based programs and model-based calculation/decision making algorithms. Very often automation requires both hardware and software in conjunction, e.g. in path planning of robotic manipulators serving machine tools, which is the domain of interest of the current work.

Among the multitude of relevant examples, an approach in robot path planning with collision avoidance, combining end effector motion planning and joints motion planning is described in [1]. A method for the use of robotic fingers in the manipulation of cartons (modelled as kinematic mechanisms) for folding operations in packaging industry is proposed in [2]. Virtual Reality (VR) software is used for modelling robotic manufacturing cells in the early design stages [3]. An analysis of robotic painting for parts with free form surfaces is presented in [4] using CAD-based software taking into account paint thickness. Robot path planning in automotive industry is dealt with in [5], using constraints on robot poses. A more generic constraint-based path planning methodology in a CAD environment is presented in [6] and demonstrated in robotic loading/unloading of a turning center.

Intelligent decisions on robotic path planning may be based on fuzzy logic [7] and on genetic algorithms [8]. Most relevant to the current work, robot path planning software is presented in [9] based on an evolutionary algorithm which decides the most appropriate path of the robot end-effector loading, holding in position and unloading sheet metal part on a press-brake. An initial path is programmed by the user programs and automated evolution of the paths follows in real time in a machine learning approach presented in [10]. The use of neural networks in real time robot path planning in dynamic environments with collision check is analyzed in [11].

In this work, robot path planning is based on robot task analysis and consequent determination of ‘standard’ robot poses. Transition between the latter is implemented with as few simultaneous joint motions as possible, in order to simplify inverse kinematics. Parametric definition of CAD models of the press-brake, handling robot and peripheral devices as well as kinematic analysis capability are the enabling technologies.

* Corresponding author: Tel.: (+30) 210 772 1457; E-mail: vosniak@central.ntua.gr
2. Problem Definition

For each product that requires a number of bends, the manufacturing process on a press-brake is similar. For the first bend to be executed, the planar sheet metal is moved from a feeding device to the press-brake properly oriented. Before the next bend the sheet metal is moved out of the machine, its orientation is changed and it is positioned between the upper and lower rams. This cycle is repeated until all bends have taken place. After the last bend, the sheet metal part is moved to the output buffer and a new planar sheet metal is loaded. The main constituents of a typical robotic bending cell are: the press-brake, the part feeder and the robotic arm with the appropriate end effector, see Figure 1.

![Figure 1: The robotic bending cell parts (a) press-brake (b) part feeder (c) robotic arm and (d) end effector.](image)

As far as the press-brake is concerned, there is no need for a detailed specification of it and therefore a simple representation with the basic parameters is used. Only some basic dimensions need to be defined along with the orientation and distance from ground level of the bending line. Besides, the whole process task analysis is designed so as only few basic movements take place near the press-brake, in such a way that the details of the press-brake design do not interfere with the robot motion; this facilitates replacement of the press-brake by another design.

The part feeder is designed to hold in place the planar sheet metal with defined orientation and position. Moreover, it keeps constant the distance of the sheet metal to be loaded from the ground level. It can be used for a variety of products and their corresponding initial sheet profiles.

The robotic arm is one with six rotation joints, providing the ability to place the sheet metal part in any position and with any orientation in the working space of the arm. In addition, the speed and displacement specifications of the joints have been taken into consideration for the simulation process. In task analysis no more than three joints work together for each time motion step, which simplifies robotic programming within similar applications (e.g. same product family, only dimension differences, etc.)

The end effector is designed to be a hybrid magnetic/vacuum effector that can handle both metallic and non-metallic sheet materials. Its basic dimensions are also taken into consideration in the simulation process keeping in mind the variety of parts that can be handled.

The parts are placed in clearly defined positions in the bending cell, that have been chosen in such a way that the needed manipulations are reduced and/or simplified as far as possible. Moreover, it takes into consideration that all parts have to be inside the effective working volume of the robotic arm.

The shape of the planar sheet metal and of the final product are analysed to determine specific patches. The main patch is defined as the one that is not rotated during the bending process; it is usually surrounded by other patches, it is in the center of the sheet and/or is the biggest patch. Every patch that is rotated during the process is a secondary patch. When this definition does not suffice, the main patch is defined as the one with the least number of bends joining it with the patch that is located in the center of the sheet. Every patch containing an edge of the initial sheet profile is defined as a simple patch. An example is shown in Figure 2. The grasping point of the sheet is selected as the geometric centre of the main patch. The side of the grasping point may change during the bends depending on the required orientation.
3. BENDING PROCESS PLANNING AND SIMULATION

Before the robot motion simulation, it is required that the position and orientation of the robot end effector is clearly defined for each step of the simulation. The robot end effector is tied to the main patch of the sheet at the grasping point for most of the steps. The steps that correspond to bending stages are excluded, because the robotic arm releases the sheet before these start. In summary, the position and orientation of the sheet must be known for the complete simulation process. Position and orientation is mainly determined by the bending sequence [12]. Software based on genetic algorithms, developed in [13], was used to determine bending sequence in 3D space. The necessary bending process simulation software was also developed.

Bending sequence was determined in previous work using a genetic algorithm for the 2D problem (“paperclip problem”) [12], and evolved next in a fully functional bending sequence optimiser for the real 3D problem [13] which is used in this work. As far as the genetic algorithm is concerned, the software is very similar to the one for the 2D problem described in [12], the chromosome representing a possible bending sequence. For n bends, each individual of the chromosome contains all the bends and each bend just once: [ID₁ ID₂ ID₃ … IDₙ], where IDᵢ is the identification number of bend i. The genetic algorithm creates the initial population, evaluates the fitness function for each individual, uses the genetic operators to create the next population and repeats the process until a certain criterion is reached. The user can change parameters such as population size, number of generations, operator influence, etc.

The difference between the 2D and the 3D problem is in the simulation of the process which is used to evaluate the fitness function. The 3D problem not only involves collision checks and tool changes, but also different types of orientation change. In addition, in the simulation process, a decision making option has been added to determine the best machine stopper position for each bend to reduce dimension faults caused by machine inaccuracy. The tools used are omitted from the chromosome to reduce calculation time [12]. Besides, the tool used is changed only if collision occurs or the user pre-defines a certain tool for a certain bend [13].

Process simulation is used in the genetic algorithm to evaluate the fitness function for each individual, e.g. mainly to determine collisions. It is also used after the determination of the bending sequence for the calculation of the position and orientation of the grasping point of the sheet at the end of each step. The user enters geometry of the initial sheet profile in the form of a complex matrix, each sub-matrix containing the geometry of every patch of the initial sheet profile. Then, each bend that takes place is defined by the patches the delimit it, as well as by bending angle and pre-defined tool used. The algorithm simulates the movement of the sheet as bending takes place, with a step defined by the user; usually 10° are enough for most problems. For each of these discreet angles collisions are automatically determined between tool-sheet, machine-sheet and different parts of the sheet, see Figure 3.
4. Handling Steps

The whole process of robotic handling of the sheet is separated into handling steps, see Figure 4. These steps are simple in that they usually require no more than three joints to be simultaneously active for the corresponding motion to be executed. Moreover, the trajectory in each of these steps is parametrically defined so as to be easy to change from one product to another. For each step, with the use of the process planning and simulation software presented in section 3, the initial and final position and orientation of the grasping point and corresponding patch respectively are known for each step. These delimit the trajectory of the end effector for each corresponding step. The robot motion between these delimiting pairs is defined using simple constraints, such as planar movement, rotation about an axis in 3D space etc. The user can change the grasping point side of the patch if needed during the process; in general, robot movement planning is an interactive process, controlled by the user.

In addition, the user can check for collision between the robotic arm and the press-brake if this is necessary using the corresponding functionality of the CAD modelling environment in which the whole process is modelled. The specific option is not default as it is time-consuming. Besides, there are relatively few ambiguous cases, for which such a collision check is meaningful. The use of the process simulation software, see section 3, already involves a collision check between the sheet and the various parts of the machine, resulting in a collision-free bending sequence. Therefore, this type of collision does not need to be checked in the simulation of the robotic arm movement.

Figure 5 shows the initial and final robot pose for each of these steps (the final pose of one step is the initial one for the next step). Some of these steps may be omitted depending on the part to be handled.

![Figure 4: Definition of robotic handling steps.](image-url)
5. ROBOT SIMULATION

The robot is simulated in a CAD environment, in this case Solidworks™ enhanced with kinematics analysis. In order to exploit in-built kinematics functions in an interactive way, the user instructs the joints of the robot to attain specific positions, corresponding to the robot pose at the end of each handling step. Joint positions are calculated using a simplified inverse kinematics model of the robot. The most complex trajectory involves end effector movement on a specific plane surface by using just three robot joints.

All other movements are simpler, involving one or two joints. In addition, for this type of movement, if one joint is used for the correction of the orientation and the remaining two for the movement of the end effector, the analysis is further simplified.

More specifically, the movement described is that of a virtual robotic arm of two joints in plane $x_0O_0y_0$ as shown in Figure 6, where $l_i$ the length of link $i$, $q_i$ the rotation of joint $i$, $(p_x,p_y)$ the position and $\theta$ the orientation of the end effector. The calculation is shown in the Appendix.
6. CASE STUDY

In Figure 7 a sample final part and its initial sheet profile are shown. In addition, the grasping point is chosen as the geometric center of the main patch.

The bending sequence determined by the genetic algorithm is 3 4 5 1 2. Using the process simulation algorithm, the user calculates the initial and final position and orientation of the grasping point and then with the use of the robot kinematic model, calculates the rotation of each joint for every step. In Figure 8, one particular step is shown for the first bend. In Figure 9 joint angular displacement is depicted for that step involving joints 2, 3 and 5.
7. CONCLUSIONS AND FUTURE WORK

The main idea that was introduced in this work in order to program robotic movements in replacement of the human operator of a CNC press brake is to break down the whole handling process into steps, each of which is simple enough to be performed by simultaneous movement of one to three robot joints. This, first, simplifies the inverse kinematics model that underlies these movements and, second, enables easy modification of each step trajectory to accommodate changes in product shape and size and the corresponding changes in number of bends. So, robustness is achieved at the expense of possibly lengthier paths.

This kind of ‘standardisation’ of robotic handling steps requires a process plan expressed as a series of positions and orientations of the product undergoing bending in consecutive stages. An advanced optimal process plan generator module was used for this purpose, ensuring collision avoidance of the sheet as well as maximum dimensional accuracy of the product.

A stand-alone graphics supported simulator of the bending steps and corresponding configurations of the product was also used in order to determine the position and orientation of the handling tool (robot end-effector) at the end of bending steps which coincide with the start of respective handling steps.

Robotic moves are defined interactively by the user in a CAD environment based on the information furnished by the process simulator and the handling steps definitions flowchart. Any modern CAD environment provides kinematics and collision detection functionality which are necessary in this interactive task.

Parts of the robot programming process can be automated, the main obstacle currently being the decisions to be made as to modifications of the handling roadmap in cases where collisions are found as a result of following some handling step ‘by the book’ when some of the explicit or implicit constraints associated with it are violated. Further work is under way in this direction.

ACKNOWLEDGEMENTS

ELKEDE is gratefully acknowledged for partly financially supporting this work.
REFERENCES


APPENDIX

\[ p_x = l_1 \cdot \cos(q_1) + l_2 \cdot \cos(q_1 + q_2) \]
\[ p_y = l_1 \cdot \sin(q_1) + l_2 \cdot \sin(q_1 + q_2) \]
\[ \theta = q_1 + q_2 \]

Moreover:

\[ ((p_x)^2 + (p_y)^2) = l_1^2 + l_2^2 + 2l_1l_2c_{q_1+q_2} + l_2^2s_{q_1} - 2l_1l_2s_{q_1+q_2} \Rightarrow \]
\[ ((p_x)^2 + (p_y)^2) = l_1^2 + l_2^2 + 2l_1l_2c_{q_1+q_2} + s_{q_1}s_{q_2} \Rightarrow \]
\[ q_2 = \pm \arccos \left( \frac{(p_x)^2 + (p_y)^2 - l_1^2 - l_2^2}{2l_1l_2} \right) \]

As for \( q_1 \), after some mathematical manipulation, the :

\[ q_1 = \arctan \left( \frac{p_y}{p_x} \right) - \arcsin \left( \frac{l_2s_{q_1}}{\sqrt{(p_x)^2 + (p_y)^2}} \right) \]

where \( c_i = \cos(q_i) \), \( s_i = \sin(q_i) \), \( c_{ij} = \cos(q_i + q_j) \) and \( s_{ij} = \sin(q_i + q_j) \).
THE DIGITAL FACTORY CONCEPT FROM MANUFACTURING TO SERVICES ENVIRONMENT

Fabio FRUGGIERO\textsuperscript{1*}, Alfredo LAMBIASE\textsuperscript{2}

\textsuperscript{1} Dipartimento di Ingegneria e Fisica dell’Ambiente
Università della Basilicata
C.da Macchia Romana, via Ateneo Lucano, 85100
Potenza (Pz) - ITALY

\textsuperscript{2} Department of Mechanical Engineering
University of Salerno
via Ponte don Melillo 1, 84084
Fisciano (SA) - ITALY

ABSTRACT

To significantly reduce development time and costs, concurrent work and change iterations (in reverse engineering approach) between design and manufacturing need to be evaluated and managed in near real time. Layout, workspace assessment and procedures, process analysis and validation can be analysed through the integration of multi-objective mathematical procedures, discrete event simulated environment and digital models (i.e., Digital Factory - DF approach). Across a digital world is possible to visualize and bear out and interplay all elements of the “future” growing factory/environment. Furthermore, brainstorms can be found on visual manifestation of hypothetical virtual environments. Expert analyst, from different fields and knowledge (like engineers and doctors), can eventually cooperate and meditate about optimization of procedures and processes on such a digital model. What is more is the possibility of resolving all design mistakes in a fictitious world (i.e., Virtual Reality environment). This significantly reduces the cumulative cost of change and time-to-market while improving quality of supply services.

In this paper, a Digital Factory approach was implemented to analyze and optimize manufacturing as well as services environment. A continuous loop between virtualization and optimization has been engaged. The carried out cases study are the muffler assembly station of an Italian automotive company and the Emergency Department behaviour of an Ireland hospital. The visualization step was, in both cases, the system validation and the final optimization phase of a more complex discrete event simulated environment. Numerical outcomes of the DF approach will be manifested. The outputs will be produced in terms of layout and facilities improvements, space usability, product/service level and quality, costs reduction and profitability of investments. Concepts like ergonomics, workplace arrangement, flow characterization, paths analysis, human/quality/cost design will be exhaustively treated and prospected. Snapshots of the digital world will be portrayed.

KEYWORDS: Virtual Reality, Digital Factory, Optimization, Manufacturing Environment, Service Industries.

1. INTRODUCTION

In our global marketplace, customers could easily look around in search of better product and services. Factory design, systems analysis and optimization, with shortening of cycle demand, compel designers, engineers and managers to make use of efficient and visionary approaches in the product and process planning phase. A correct and careful planning process, aimed at customer expectation/satisfaction, represents an opportunity to future improvements and benefits. To improve planning results and to carry out customer oriented products/services, and effective production systems, is necessary a reconfigurable and concurrent engineering approach in product, factory design, system configuration [1]. This takes place when all people, who collaborate for the planning, place simultaneously their knowledge at project's disposal [2]. In order to have quality and productiveness, the persons responsible of planning, production and maintenance need to cooperate with the product design group in product’s definition phase. They have to communicate through a common interface [3].

An error into the planning phase manifests much more costs than a “quality management” approach. It is needed most attention to the planning activities that mutually affect the product quality, production process. Anyway, the

\* Fabio Fruggiero: Tel.: (+39) 0971 205196; Fax: (+39) 0971-205160; E-mail: fabio.fruggiero@unibas.it
best way to face up to the gaps is to meet design flimsiness at the early stage of the product lifecycle statement. In the early design stages, the impact of a design change could be minimal because few process plans have been defined and no resources committed. The “Rule of 10” could be applied: the cost of correcting a hundred-dollar mistake in the design phase could increase exponentially across prototyping, setup and production phases to one million dollars after the product has reached the market [4].

The planning phase has generally two main targets: the output, in the form of goods and services, and supply processes. To develop an efficient system, it is necessary to interactively consider the impact of design changes in the plant and process in such a way that the physical manufacturing of the product/service meets quality, time, cost goals and customers satisfaction. Moreover, the traditionally concept of planning phase is as time-consuming (e.g., the product is late on markets), myopic (e.g., mainly looking at product and not process design), out of enterprise context (e.g., poor of customer’s ratings). To overcome these gaps and to maintain the results as planning, the physical model optimization need to be performed. “Virtualizing” the system is a rewarding approach [5]. In the digital environment, alternative process design configurations, in little efforts and time, can be re-engineered and, then, evaluated in a concurrent manufacturing extent of the virtual world (Virtual Manufacturing - VM) [6]. The use of tools and methodologies of digital/virtual prototyping (i.e., Virtual Reality - VR systems) can make easier the system optimization process. Across a computer supported world is possible to visualize and bear out and interplay all elements of the “future” growing factory/environment. This approach gives manufacturing planners early access to in-work designs to allow for concurrent design manufacturing planning and ongoing manufacturing input into design decisions [7]. They can interactively change the layout, resources, machine features, and observe how the system immediately responds [8]. Brainstormings can be founded on visual manifestation of hypothetical arrangements. The bent is leading to deem only after a successful virtual validation the system’s design is implemented and/or modified in the real factory/environment. The 3D visualization of environment is going to be the final output of a more complex optimization procedure [9]. The aim will be a taxonomic rise in workplace ergonomics/pleasantness, workplace/worker safety, increase in productivity, client satisfaction and product/service quality going up overall cost reduction.

In manufacturing, as well as into the Health care management, the digital representation act as a medium to the optimization. In this paper a looped approach between system’s configuration and simulated virtual model has been engaged. The application of digital tools to optimization, from a flow shop manufacturing enterprise to a complex “job shop” service system, are the key elements of the implementations. This paper exhibits the process of developing the optimal layout design and system configuration, and flow analysis by providing the management with multiple layout configurations and showing the impact, basing on simulation output and virtual manifestation, of the arrangement at each stage. It is mainly divided in three section. The first one is dedicated to the Digi ta Factory concept and the definition of optimization process. Firtsly, the authors described the implemented procedure to the optimization and subsequently they spent time into description of the implemented cases. A different section is dedicated for each real implemented settings. The outputs will be described in terms of productivity of capital stocks, reduction of Not Value Added Actions, design flimsiness, quality of services, system lead time, space and usability and accessibility, perception of the labor. Considerations on benefits of such an approach end this work.

2. THE PROCESS POINT OF VIEW IN THE DIGITAL FACTORY APPROACH

To implement a Total Quality Management (TQM) system, a Total Industrial Engineering (TIE) have to be made up by optimizing the supply processes with a consequential rationalization and standardization of human employment and materials handling and flow paths and services [10]. The integration of different tools in digital frame has been implementing into the filed of manufacturing for few years [11]-[12]. This overture has been mainly noted as Digital Factory (i.e., DF). The DF accomplishes the TIE purposes [13].

Digital Factory is the new overture in production systems that “substitutes moving virtual bits for moving physical atoms” [14]. DF is a generic term for a comprehensive network of digital models, methods and tools [15]. It includes product and components characterization and design, data analysis blocks, simulation (e.g., by discrete event models) and optimization procedures, creative problem solving methodologies and virtual ramp up (fig. 1). These are entirely integrated with an upper data management system. DF allows to analyze all aspects that can be observed in a real factory, providing higher insight about the processes, material, fluxes and even human activities and ergonomics [16]. Across the DF approach there is a methodology for organizing the information to the development of the virtual product (i.e., design) and the virtual factory (i.e., planning). The approach links digital design, process and production requirements, and provides a lifecycle view of the complete product definition. The integration of product/process development reduces the system lead-time while improving the process productivity
and quality (of product and process and services) and reducing the overall costs. Its overture facilitates intense collaboration between design and manufacturing planning, from different knowledge background (e.g., engineers and doctors) [17]. Moreover, this approach let them to evaluate designs and make changes, incurring foreseen virtual costs and business risk, quickly and accurately in the virtual development phases when the cost of change is low [18].

Figure 1: The bulks of information in Digital Factory behaviour

A visual optimized approach causes systems designers to think about certain significant issues (e.g., system control and production rise and costumer satisfaction) long before they normally would. The common view of DF framework is as featured by five components fully integrated through a common database for the coordinated management of information generated [19]. Modules for external relationship need to be implemented. This phase is performed by an I/O relationship network pointing out functionalities existing among the element of network. This set the boundaries and the frame of the case in issue. The model structure, the visual ground, the schematization of production system and tasks, the flow and layout analysis are the input data of any modeling phase. In this stage the Product, Process and Resource (i.e., PPR) and Bill Of Material (i.e., BOM) have to be defined. PERT (i.e., Program Evaluation and Review Techniques) and CPM (i.e., Critical Path Methods) are generally used into the modeling phase. Furthermore, the understanding of the probabilistic evolution of the system gives the possibility to implement a consistent model. A statistical data analysis has to be performed in order to depict and bound and prospect the range of alternatives (e.g., a macro identification of trends and possible bottlenecks need to be carried out). The digital simulated implementation accomplishes a complete evaluation of alternatives. A simulated environment is able to dynamically appraising the functioning of the system while evaluating detailed performances for operations choices (i.e., PPR amount). An optimization approach needs to be fixed. It is always defined in order to have a huge database (fulfilling constraints and tabu sets) of possible alternatives. The exploration, while simultaneously exploitation of domain, should be bound up with the definition of a “general purpose” optimization function (mainly based on a balancing assessment) and creative problem solving approach [20]. Modules of final visualization of the system, in 3D digital world, give the possibility to prospect changes and evaluate the results. This approach provides a robust yet flexible configuration management system.

2.1. **THE DIGITAL FACTORY AS A COMPREHENSIVE TOOL FOR TO THE SYSTEM OPTIMIZATION**

As concurrent engineering increases and production cycles shrink, development teams and the systems that support them have to manage a stream of real-time design changes, each requiring routing and reconciliation. In a new view of quality management approach it is possible to substitute the evaluation phase of improvement with a virtual optimization process. The digital 3D world becomes the valuable aspect when clustered units (in knowledge and experience) have to cooperate. Shunting a connection by the Deming (i.e., a cycle of Plan-Do-Check-Act) principles [21], the testing phase of the planning improvements can be evaluated into the virtual model without/before changing the real assets. The procedure of quality improvement can be performed across a cycle of Plan-Virtualize-Check&Optimize-Action. The cycle can be stated as: plan the system, set up a model and individualize problem solution; build the complete model of manufacturing frame (i.e., DF mock-up) in virtual environment; evaluate performances and optimize according to analytic/expert procedures and simulated settings; translate the virtual hub in real world. The advantage is to implement the tasks only after the successful of an
optimization procedure. Moreover, it is just possible to test the solutions which fulfill the requirements of the optimization loop. According with the virtual hub is possible to identify the optimal arrangements and predict results while performing system’s analysis in “fictitious” environment (fig. 2). The understanding change within the context of lifecycle processes and resources as the product matures is the competitive advantage of a digital factory approach.

Starting by a plethora of entry data, the identification of the main features of the model through a statistical assessment phase (shape and modal values need definition) can be accomplished. The representation of the system according with flow charts and mainstream parameters helps to validate the mock-up plans in expert insights. Analyst, managers and engineers have the possibility to validate and modify procedures from different points of view. The optimization of design and plant revamping implies assuming certain compressor conditions to optimize the process. An optimal condition is usually found that optimizes the balance between process turbulence and unjoin unpleasant surprises [22]. A discrete event simulation model, mainly based on meta-heuristic procedures, user defined in relation to the particular problem in issues, helps to analytical evaluate the specified optional configurations [23]. The final optimal arrangement is obtained only after a procedure of visualization in virtual environment where is possible to introduce parameters like space usability, ergonomics, tool handling, maintenances requirements, workstation configurations, tool handling, safety, flow flimsiness and “jam” etc… In this super heated environment, alternatives scenarios could become increasingly diverse and unpredictable. Maintenance may request a change to make it easier to repair; while shop floor staff might ask the manufacturing planner to shift work from one control station to another to improve production flow. Moreover, it is needed to evaluate how human workers and plant resources are impacted by post-production changes. Only visualization structure yield positive results in these cases [24]. The model can be analyzed, and eventually simulated, stating a virtual production mock-up given the available time, space, resources and pointing out alternatives. A decision making process (e.g., Analytic Hierarchy and Network Processes characterization [25]) can be eventually embedded into the event driven simulator. Before comparing alternatives among each other, each scenario is deeply optimized with preliminary runs of the model, that help defining the most appropriate technical configuration, like for example the best worker/staff arrangement, the optimal rota and layout. The 3D model manifests the visual report/output of the implemented optimization process.

2.1.1. The manufacturing environment optimization: A DF investigation into muffler assembly station

A worldwide automotive plant, which produces approximately 250,000 vehicles (in different models and configurations) per year, was analyzed and optimized by the mean of DF approach (table 1). The plant is subdivided in ETUs (Elementary Technology Units); the area under analysis in this paper is the muffler assembly station. The job performed in this ETU, consists of the following operations: exhaust pipe preparation, exhaust pipe assembly, axle assembly, stirrup on hub hand, axle fixing, fixing muffler pipe on the vehicle bodywork, fixing front crossbar, quality control of ETU performances. The ETU is attended by 14 operators, working on three shifts, which produce up to 220 vehicles/day. Statistical data analysis, evaluation of flow paths, safety and quality considerations, cost assessment have been used into the optimization approach [26].

Figure 2: The event flow chart for the process of optimization across the Digital Factory approach
Manual and automated tasks have been simulated into the virtual environment aiming to improve system performance, workspace and worker environment, product quality. In order to investigate new plant configurations and performances, in terms of balanced utilization of resources and saturation rates and service levels to customers, virtual representation/validation have been implemented (fig. 3). Anyway, a database of collected data and an expert analysis identified as the most critical production tasks the catalyst, muffler pipe assembly task and the assembly of muffler pipes on the vehicle. These are going to be the area/jobs/tasks under analysis (fig. 3c). Two or three operators (according to production rate) attend these phases. A worker prepares the catalyst, the pipe (composed by the catalyst, the temperature probe, the metallic gasket and the stirrup) and finally he installs the muffler pipe on the vehicles. The responsible for the muffler pipe assembly on the vehicle brings the completed pipe from the catalyst preparation area and leaves it on a conveyor system; then, the second operator brings the axle from the vessel on the line edge to the conveyor. Subsequently, both operators assemble the break pipes, the ABS sensors and the wire of the hand break (the muffler pipe is mounted on the vehicle in the next station). A conveyor puts the item on the assembly desk. A critical overview to the workplace arrangement and tools handling in simulated virtual environment is able to underline some desirable logistic improvements. These elements can be introduced into the optimization process across to DF mock up. The traditional methods, to ergonomic study, require large amounts of times, due to the subdivision of tasks in basic actions, the evaluation of the key positions (e.g., angles, distances, etc...) and the calculation of the risk indexes [27]. On the other hand, simulated environment and digital dummies, with bio-fidelic kinematics and advanced anthropometric scaling, show how to organize workers’ environment, to design tools, to perform operations and temporize and standardize the tasks [28]. The digital implementation with dummies allows the visual evaluation of not-ergonomic postures, tools, facilities and performs a Motion Time Measurement (i.e., MTM) optimization (fig. 3b). The calculation of NIOSH Lifting Index (i.e., 1.41, 1.46, 1.54 for each operator) [29] and Rapid Upper Limb Assessment (frequency capability near the 60% and mean RULA index 4 as acceptable stress) [30], with percentile and anthropometrical considerations, underlines the validity of tasks in the optimized arrangement. Planning a new configuration of the vessel, workflow and workflow according to the virtual mock-up of the implemented simulated environment carry system to save the 33% of the transport costs (e.g., arranging three rows of vessels on trucks rather than two on initial configuration), with a small investment of €42,150 and a payback period of 8 months (fig. 3a). A Not Value Added Action, performed across the evaluation of different alternatives into the virtual simulated frame gained the reduction of 52%, compared to the original layout, of the distance covered by the operator during workshift. An exhaustive simulation loop prospected the possibility of criticalities for the operator’s safety (i.e., in the procedure of handling the muffler pipe from the assembly bench to the in-line vessel). To avoid all possible contacts of the operator with the muffler pipe, a protection (metallic net) is used; however, it holds over a possible risk for the operators that could get close to the assembly bench. The 3D model of the systems was developed in order to verify in deep details, and manifest in results, optimal paths and construction sequences and assembly solutions to achieve the maximum safety, space and work efficiency. A more focused production structure has been consequently carried out with competitive costs advantages.

Table 1: Summary results of the DF optimization approach into manufacturing behaviour

<table>
<thead>
<tr>
<th></th>
<th>Layout track length</th>
<th>Length covered each shift</th>
<th>Ergonomic analysis (NIOSH evaluation)</th>
<th>Quality &amp; safety</th>
<th>Material handlers [€/Year with 220 items/24h]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BASE CASE:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLD methods, 3operators</td>
<td>28[m]</td>
<td>2309[m]</td>
<td>1828[m]</td>
<td>Operator 1</td>
<td>Operator 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>540[m]</td>
<td></td>
<td>1.41</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>25%</td>
<td>66.9%</td>
<td>19.7%</td>
<td>//</td>
<td>//</td>
</tr>
<tr>
<td><strong>NEW METHOD:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF opt, 2operators</td>
<td>21[m]</td>
<td>764[m]</td>
<td>1467[m]</td>
<td>Operator 1</td>
<td>Operator 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.02</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>25%</td>
<td>66.9%</td>
<td>19.7%</td>
<td>//</td>
<td>//</td>
</tr>
</tbody>
</table>

Reduction (% of base) 25% 66.9% 19.7% // 27.7% 19.2 // improved 33.3%
2.1.2. FROM MANUFACTURING TO SERVICES: THE DF IMPLEMENTATION INTO AN EMERGENCY DEPARTMENT

The care of patients and quality of services is the main objective of the health care system. Emergency Department (i.e., ED) performance concerns with resource amount and allocations, process standardization, patient’s classification and prioritisation, patient’s waiting and throughput time, queuing disciplines. The objective of this paragraph is to present the application and usefulness of DF approach in a process of optimization of ED services and resources. Across the general approach to the optimization (as described in previous paragraphs) the authors are going to present an extended way-out into digital tools. The lack of standardization into services environment was the main difficulty into the implementation of such a DF model [31].

Figure 4: The Virtual Emergency Department environment in different snapshots.
From the Discrete event simulation view to the 3D virtual environment.

A real Irish Emergency Department has been investigated. Flow paths validation, layout assessment, space usability and procedures’ validation, cost’s estimation and scheduling procedures in alternative configurations/arrangements have been investigated through the implementation DF approach. Modelling the operation (treatments) is very intuitive in DF approach: user can view and prospect each plant workstation and configuration individually in the simulated 3D environment. A user interface has been designed to insert treatment time and to join real and virtual flowing. The optimisation of variables (patient’s throughput time, and number of resources and costs) has been realized and visualized across an optimization routine based on the minimization of a user’s defined function (taking into consideration hospital efficiency and efficacy) [32]. Found on the implementation of a more complex simulation model, pointed on job shop rules, the virtual representation aims to be the validation steps of hypothetical alternatives/configurations and layout in the process of supplying services to patients. The discrete event simulation model tests the system configurations in terms of throughput and waiting time based on defined rules and procedures (the mean patient Waiting Time was evaluated in different configuration, i.e. opt case 214 min, with the characterization of waiting time for first visit, depending on system overcrowding, i.e. Reaction Time as opt 65min, and total Time of Stay into the ED facilities, i.e. opt case 265). While the model is running, the digital manifestation of the environment supports more than a validation opportunity. Moreover, the competitiveness of each scenario strongly depends on conditions not well defined, like from the patient arrival rate, triage level, resource availability. The optimization of the system is mainly based on when the treatment will be available and for how long its supply will be guaranteed. So a correct statistical module and a robust planning model become of fundamental success in such an implementation. The arrival rate (in categories and pathologies) has been “Poissonly” distributed with different shape parameters according to real registration desk data (i.e., the average event in the simulated slot time).

The layout of ED in CUH was investigated in order to improve system’s performance. An analytic optimization procedure based on Quadratic Assignment Process – QAP with heuristic approach [33] represents the implemented mathematical formulation to validate the location of a set of treatments room. The physical facilities of ED in CUH consist of a registration desk (RR), a waiting room (WaitR), a triage room (T), an X-ray area, 12 treatment rooms for major patients (57% of arrival) and 4 beds for minors treatments (32% of arrivals are categorized as minor), 4 resuscitation areas (1.4% of incoming patients need its facilities) and separate stations for the unit secretary, nurses, nurses aides and physicians. The ED treatment zones were switched, in location, considering flow paths and zoning constraints. The flow of patients $f_{ij}$ from facility $i$ to facility $j$, the distance $d_{kl}$ between location $k$ and location $l$, and a weight account (i.e., a function of assignment relevance according to expert analysis) $b_{ik}$ of placing facility $i$ in location $k$ have been analyzed in a combinatorial approach in order to defined a fitness function of allocation. Improvement has been prospected in number of cubicles, mainly for major patients which represents the 57% of arrivals (table 2). The data were modeled as according to the entry registration data. A tabu list forbade the assignment to a particular location (e.g., the location of resuscitation close to ambulance’s entrance). Mathematically the problem is the allocation of a set of facilities to a set of locations [34]. A Simulated Annealing (SA) procedure was implemented [35]. The simulation model permits to evaluate the current process behaviour and to predict the
impact of changes (fig. 4). The scheduling of tasks into the ED facilities were investigated according to a balancing approach with meta-heuristic exploration and exploitation rules. In such an approach the authors considered the impact of different system’s configuration evaluating the number $N$ of resources into the ED system, the workload $W_i$ of resources $k$, the Length of Stay of patient $i$, the Saturation rate between interchangeable, and no, resources. The processing time has been modeled according to Fuzzy rules [36] as after doctors approval. The line balancing approach detects fundamental problems (unnecessary and/or duplicated activities), evaluates the sequence of operation into the department in issue (analysing scheduling process). The authors defined the optimal scheduling in relation to user targets (optimization of patient’s scheduling, patient’s throughput time), and optimal configuration of resources (i.e., in opt case: 8 Nurses, 5 Porters, 3 Registration desk operators, 2 X-Ray technicians, 7 junior doctors - SHO, 6 Register, 1 Specialist called for consulting – see results in table 2).

The advantages of having a 3D-DF approach/model is that while performing a visual discrete event simulation analysis of the flow paths and configurations the managers are able to evaluate the procedures and direct manage the queues. Moreover, the visualization of emergency facilities is able to give users an interplay procedure in the process of fixing main rules and procedures. Inquiries between pathologies (e.g., the testing of the system in case of earthquakes), bed allocation rules (e.g., changing the availability in specific divisions as surgery and/or cardiology), space usage (e.g., testing the waiting areas) and layout opportunities (e.g., evaluating the internal areas for future expansions) have been implemented based on the digital design of the system. The managers impacted directly with proposals and alternative insights. Furthermore the ED facilities were inquired without touching the real services streaming. The quality of the system was assessed in consideration of service level to patients.

Table 2: Summary results of the DF optimization approach into ED behaviour

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Of</td>
<td>Wait Time</td>
<td>Reaction Time</td>
<td>Lay-Out</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stay [min]</td>
<td>[min]</td>
<td>[min]</td>
<td>[units]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE CASE</td>
<td>13 [unit] 65%</td>
<td>2 [unit] 50%</td>
<td>3 [unit] 65%</td>
<td>1 [unit] 65%</td>
<td>6 [unit] 50%</td>
<td>4 [unit] 35%</td>
<td>2 [unit] 45%</td>
</tr>
<tr>
<td>NEW METHOD: DF OPT</td>
<td>8 [unit] 75%</td>
<td>5 [unit] 60%</td>
<td>3 [unit] 67%</td>
<td>2 [unit] 70%</td>
<td>7 [unit] 75%</td>
<td>6 [unit] 65%</td>
<td>1 [unit] 78%</td>
</tr>
<tr>
<td>IMPROVEMENT</td>
<td>.5 [unit] 15%</td>
<td>+3 [unit] 20%</td>
<td>// [unit] 3%</td>
<td>+1 [unit] 7.6%</td>
<td>+1 [unit] 50%</td>
<td>+2 [unit] 85%</td>
<td>+1 [unit] 73%</td>
</tr>
</tbody>
</table>

3. CONCLUSIONS

The aims of the present paper is to manifest components, application fields and results of the Digital Factory approach to system optimization. The integration of different tools in digital frame has been implementing into the field of manufacturing for few years. Moreover, the DF approach remains enclosed to the industrial environments. Presenting two different cases study, without going into details of results and descriptions (to refer to bibliography for their characterization [24-30-32]), the authors provided how the application of Digital Factory can be exhaustively extended from automotive industrial environment to health care fields. In the present paper the Digital Factory has been characterized in its components. The DF approach has been described in its basic components: modeling, statistical, simulated, optimization and visualized frame. If integrated in a common database of continuous loop they provides a robust yet flexible configuration management system that allows changes to be effective for design only, manufacturing only, or both. In manufacturing as well as in ED services environment, the employment of DF techniques has allowed to optimize layout configuration and its defaults, inefficiencies, product/service quality and costs, space usability, service to patient. Furthermore, in visual environments it is easily the process of balancing the operator’s tasks and rearranging the workshop’s layout. Besides, the imagination of future assessments can be greatly furthered by the creative management of visualization in Virtual Environment.

The authors pointed out and provided for how the application of Digital Factory approach can be exhaustively extended from manufacturing environments to health care fields into a system optimization perspective.

REFERENCES


Object techniques in an integration of technical and organizational production preparation processes

Cezary Grabowik

Institute of Engineering Processes Automation and Integrated Manufacturing Systems
The Silesian University of Technology
Akademicka 18A, 44-100 Gliwice, Poland

ABSTRACT

The paper presents an approach to a problem of design features decomposition process. Manual decomposition of design features is described. Based on the result of a decomposition process an open structure of design features was worked out. This structure was a base for programming activities. The open structure was implemented with a GRIP language in NX 6.0 system. Thanks to this program modeling of axisymmetric products such as shafts, sleeves and gear wheels is possible. The result of a modeling process is a product model which is recorded as NX 6.0 part and additional by means of the two text files. These files are used for a representation of a product structure, dimensional and tolerance requirements. The open structure of design features was also a base for applying of object programming in order to prepare a CAPP system. In this case an object structure of classes that represents a product model in a CAPP system environment is used.

1. INTRODUCTION

From a definition a technical production preparation consists of a constructional production preparation and a technological process preparation. The constructional process preparation rely on designing of a complete new product or making some modification in an existing product whilst technological production preparation rely on planning of manufacturing processes. A manufacturing process planning is beyond any doubt one of the most important activity in a domain of the technical production preparation. The main goal to reach during this phase of the technical production preparation is to define the whole structure of a manufacturing process, that is a sequence of technological cuts and operations which have to be performed in order to transform a semi-finished product into a product which performs all necessary conditions in relation to: dimensional accuracy, shape accuracy, surface quality. This manufacturing process should also guarantee the minimal manufacturing cost and the minimal labour consumption taking into account available manufacturing resources. There are many methods available for automation of a manufacturing process planning for example: group technology, manufacturing process typification and first of all systems of computer aided process planning – CAPP systems. CAPP systems are usually built with variant, semigenerative or generative method. The variant method is based on the two ideas, the idea of elements/products classification by Mitrofanow/Opitz and the group technology. This kind of a CAPP system is built in the two stages. The first stage is a preparatory stage. During this stage the following tasks are being done: elaboration of a method of elements classification (nowadays the most popular are those which are based on AI techniques application e.g.: neural nets), creation of part/elements families, elaboration of a manufacturing process model for each part family, record of worked out manufacturing processes into a technological database. The second stage is an operating stage during which a new code for a new product is prepared. The product is next classified into a single family. A model of manufacturing process for chosen family is modified in order to adopt it to a product technical description (during the adopting process some of manufacturing operations are modified, some are removed or added). The main disadvantage of the variant method is that it is quite similar to manual manufacturing process planning. It means that during preparatory stage the presence of an expert is needed in order to prepare models of manufacturing processes next stored in the technological database. A generative method and a semi-generative method are alike. The difference between them is slight. In the case of a semi-generative method changing of manufacturing process model is possible. A user can also get unfinished manufacturing process just from a CAPP system and next complete it in manual way. Planning of manufacturing process with a generative CAPP system is based on following assumptions: a product is described by means of a finite set of design features,
for each design feature which has to be manufactured a set of manufacturing variants is worked out. A set of manufacturing variants is a base for making a complete manufacturing process.

1.1. CAD/CAM INTEGRATION

The most critical chain in a data exchange process is an information link between CAD, CAM and CAPP systems. Nowadays these systems act rather in the way of isolated automation islands [2, 3, 5, 8]. It causes that an engineer has not appropriate knowledge about overall manufacturing cost at the design stage. Therefore it can also have a strong influence on a final project economic efficiency. In this case very often appears a question who should be a user of large integrated CAD/CAM/CAE systems such as for example NX 6.0, CATIA etc. in a domain of NC machines programming. Should it be a designer who has certain amount of a technological knowledge? Maybe these actions should be done by a process engineer? This question comes from the fact that in many cases they talk about the integration between CAD, CAM systems but they forget to answer a question what is a source of manufacturing processes plan, what does a manufacturing process plan come from?

The answer to this question is to introduce a CAPP system as an integrator element in a process of data exchange between CAD and CAM systems. The proposed solution in figure 1 is shown. The analysis possibility of potential manufacturing costs at the design stage is an additional advantage of this solution. There is a weak point of this solution, these systems usually work in a batch manner (they usually play a postprocessor role) it means that they begin work as soon as possible but always after finishing the design stage. It unfortunately corresponds to a sequential model of action in a design manufacturing chain. This luck of collaboration between these systems results directly from the lack of the bidirectional communication between CAD/CAPP systems. Therefore is very difficult to implement the concurrent engineering philosophy but there is one solution which can be used in order to solve this problem. This solution will be presented further.

![Diagram of CAD/CAM/CAPP System](image)

Figure 1: The CAPP system as the integrator element

2. APPLICATION OF OBJECTS TECHNIQUES IN TECHNICAL–ORGANIZATIONAL PRODUCTION PREPARATION

The characteristic feature of an object technique is that its organization is similar to the human world perception. The object technique allows to connect in a particular object compound data structures with procedures which operate on its internal structure. Procedures are also responsible for providing communication between objects. Taking into account requirements which are set during a CAPP system creation process the main advantages of object techniques are as follows [2]:

- objects allow to describe the way of expert thinking and to record his own knowledge,
- objects enable to represent relations, each relation can be also treated as an object,
- classes and class objects allow to describe processes which proceed in real systems.

Term of objects techniques in the case of their application in the process of computer aided technical-organizational production preparation is considered dually. It means that author on the one hand by term of object techniques mean: design, technological and functional features and one the other hand mean OOT techniques such as: object oriented analysis, design and programming. In the first case features techniques is used for a product design representation, manufacturing technology representation. A product is represented by means of a set of design features. Each particular design feature represents particular distinguished feature of a product. In the case of
application of design features for a design representation, selection of the method of building of constructional database is needed. The constructional database contains information about prototypes of particular objects. Nowadays the most common used methods for constructional database building are [4, 5, 6, 8, 12]:

- methods of feature recognition,
- methods of automatic features decomposition,
- the manual method of an arbitrary decomposition of features.

In the presented work the third method of constructional database building was employed for the sake of its advantages such as: possibility of participation of experts in a process of features decomposition (this guarantees proper degree of database representative), the greater speed of the decomposition process in comparison with other methods, implementation and database developing ease. This method is appropriate for a database building for each kind of products but the scope of the present work is limited to the group of axisymmetric products such as shafts, sleeves and gear wheels. This limitation was made because for this group of products a feature identification process is simpler in comparison with other groups of products. It comes from the fact that for a group of frames for example only a part of surfaces is treated while the rest of surfaces remain in their previous state which comes from the previous manufacturing processes (i.e. cast). In the case of axisymmetric parts in principle, apart of semi-finished product all product surfaces are treated. In figure 2 we can see a model of a shaft and design features which were manually decomposed from the shaft model by an expert.

An expert during the decomposition process distinguished in the shaft model the following design features: pins marked out with symbols CZW1÷4, chamfers FZZ1÷2, fillets PPR1÷2, center holes NAK1÷2 and one design feature WW1 which denotes splines (note, all abbreviations come from Polish if you want to write them in English you can call them for example CLP1 which means cylindrical pin no. 1, CHO1 – center hole 1 etc.).
A design modeling process with the design features method use can be realized in the following three ways:

- the incremental method – the constructional method, in this case a product model is created step by step by adding following features to each other. It corresponds to a rapid manufacturing process.

- the removal method – the technological method, a product model is created by removing following constructional features from a semi-finished product. For axisymmetric products it corresponds to removal process that is realized in the real world on a turning machine;

- the mixed method – the constructional-technological method, a product model is created by adding or subtraction particular objects to each other.

From author point of view the best method for designing a CAPP module (if a CAPP module is a part of a CAD system) is the technological method for the sake of its similarity to a manufacturing process. If the technological method is implemented, CAD system plays a role of a data source about a product model in the special way. Therefore it is possible to make a user interface in such a way that allows to make a technological process in the background. Using this scenario a CAD system user makes a product model and a model structure. This model structure is nothing other the structure of a manufacturing process of a product it means the order of design features corresponds to a sequence of manufacturing operation and cuts. In this case a CAPP system inference engine has a very simple structure. In spite of undoubted advantages of this solution it is efficiency is in principle limited to a group of axisymmetric parts such as shafts, sleeves etc. If one wants to widen a filed of system action to other group of products (frames, levers etc.) it is necessary to work out an application which works in a CAD system environment with the constructional–technological method. Moreover the paper author considers that application of the constructional-technological method from a user point of view is more friendly. For example a modeling process of shaft in this case rely on inserting following design features such as shaft pins in order to get a rough shape of a shaft. This shape of a shaft corresponds to a state of a manufacturing process after rough manufacturing. The difference between the rough shape of the shaft and the shape we get after rough manufacturing is as follows: during the modeling process we model a product without taking into consideration dimensional deviations – we model it just for nominal dimensions whilst after rough machining some dimensional deviations can stay over some shafts pins. Next the final shape of a shaft is made by subtracting design features such as: grooves, undercuts etc. The modeling process with constructional–technological method in figure 3 is shown.

![Diagram](image-url)

Figure 3: The modeling process with mixed (constructional–technological) method
Figure 4 presents the open structure of design features worked out at the present stage of the research. The identification process of design features was led taking into account the following rule, for each particular design feature a single machining cut, an alternative machining cut, group of machining cuts or alternative group of machining cuts should be ascribed. It corresponds to the situation in which a single manufacturing feature or group of manufacturing features (machining operations or cuts) corresponds to a particular design feature. This type of connections answers to relations such as one-to-one or one-to-multi. Such a definition of relations which appear between design and manufacturing features makes process of a CAPP system design and implementation much simpler.

Figure 4: The open structure of design features
Appearance of relations one-to-multi is very important from collaboration between CAPP and CAP/PPC point of view. Accessibility of this kind of relation allows to create multivariant technological routes. It is essential in the case of systems which are able to solve problems connected with scheduling and rescheduling problem taking into consideration possibility of disturbance appearance.

In figure 4 we can distinguish the two kind of lines: the dash–dot line that sets a border of classes which are used in the incremental way (constructional method) and the dot line that sets a border of classes which are used in the removal way (technological method). In the mixed method both type of classes are used. The mixed method makes possible to widen a CAPP system action filed to a group of products which do not belong to the group of axisymmetric parts. The open structure (see figure 4) was base for programming activities. This structure was implemented into two different environments. In CAD system it was implemented with GRIP language whilst in the CAPP system it was implemented in Borland Developer Studio 2006.

The GRIP language is a procedural programming language therefore it does not support the object paradigm. In order to simulate object behavior for a set of design features, structure of each design feature was written in a separate program module. It means that the number of modules has to be equal to the number of design features excluding modules responsible for a database management and other activities which are not connected with design representation. Each program module corresponds to a structure of a class in the object paradigm. Of course it is not possible to make hierarchy of modules in this way as it is made in the case of classes; as we know this opportunity is the most powerful advantage of object programming languages. So, why is it worth writing this structure in separate modules instead of writing it into a one large program, the answer is very simple. When program is divided into many separate modules we can make same changes in our program by changing code of a single module so it is quite impossible to damage the rest of program code. Moreover if we make some mistakes in the code of a module the rest of the program can work without any problem. If the open structure is written in the one large program when we make same changes in its code there is a high risk of damaging our program. Recording of a design feature in the way of separate modules allows to develop program simply and quickly by adding a call to a new developed module into a management module. Worked out program is next used in the Unigraphics/NX 6.0 environment in a batch manner. This program called SHAFT–CAD consists of a management module, modules that represent particular design features as it was written above and modules responsible for a selection data from a standards database. The structure of SHAFT–CAD program in figure 5 is shown.

![Diagram of SHAFT-CAD program](image)

In figure 5 the following abbreviations were introduced: MM – a management module, DFD – a design feature database (organized as a set of separate modules), SDMM’s – management modules group responsible for a standard database servicing, SD – a standard database, NX 6.0 – environment of CAD/CAM/CAE system. A user of the SHAFT–CAD program begins its work with modeling a rough shape of a model by adding following shaft’s pins one to another. If the rough shape of the model is finished a user can add to the model features which are subtracted from the rough model so one gets a finished product model. A user of this software is supported by SDMM’s modules during a modeling process. For example, in the case of inserting a design feature that has a standard shape and dimensions (grooves, undercuts, threads, etc.) standard data are introduced into appropriate edit boxes of a user interface window automatically according to a suitable standard. We have to remember that in same cases the design feature dimensions are strictly connected with pin’s dimensions. If SDMM module cannot match feature dimensions with pin’s dimensions in accordance with a selection rule for example, an automatic selection is
impossible. In this case, as it is said in the fuzzy logic it is better to give a user an answer that is sure for 80%, for example, instead not giving him any answer. Therefore program tries to propose same data values and a user has to check and correct them. The SHAFT–CAD program plays a preprocessor role for a CAPP system so SHAFT–CAD is responsible for product data recording. A product data consists of an information about product structure – relations between particular design features used in modeling process, a shape and dimensions of design features. The information about a product structure in a text file with *.stc extension is recorded whilst a model geometry it is particular design feature dimensions in a text file with *.geo extension is recorded. This information is next used by a CAPP system in order to prepare a manufacturing process. The modeling process with SHAFT–CAD use in figure 6 is shown. A form of a model structure file and geometry file in figure 7 is shown.

Figure 6: The modeling process in the SHAFT–CAD program, a) the rough shape of a shaft, b) the finished shape of a shaft

<table>
<thead>
<tr>
<th>MODEL TOPOLOGY – (H:\G\GEOMETRIA\t5.stc)</th>
<th>MODEL GEOMETRY – (H:\G\GEOMETRIA\t5.geo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>02–16–09 -- 11:02</td>
<td>02–16–09 -- 11:02</td>
</tr>
<tr>
<td>CZW1</td>
<td>CZW1</td>
</tr>
<tr>
<td>CZW2</td>
<td>KT_OBK_PRZ</td>
</tr>
<tr>
<td>CZW3</td>
<td>1</td>
</tr>
<tr>
<td>CZW4</td>
<td>PPOCZ</td>
</tr>
<tr>
<td>CZW5</td>
<td>0.0000</td>
</tr>
<tr>
<td>CZW6</td>
<td>SR_D</td>
</tr>
<tr>
<td>CZW7</td>
<td>25.0000</td>
</tr>
<tr>
<td>CZW2 WWZ1</td>
<td>DL_CZOPU</td>
</tr>
<tr>
<td>CZW4 RWZ1</td>
<td>20.0000</td>
</tr>
<tr>
<td>CZW4 RWZ2</td>
<td></td>
</tr>
<tr>
<td>CZW4 RWZ3</td>
<td></td>
</tr>
<tr>
<td>CZW4 RWZ1</td>
<td></td>
</tr>
<tr>
<td>CZW7 FZZ1</td>
<td></td>
</tr>
<tr>
<td>CZW5 PRZ1</td>
<td></td>
</tr>
<tr>
<td>CZW6 RSZ1</td>
<td></td>
</tr>
<tr>
<td>CZW3 PRZ2</td>
<td></td>
</tr>
<tr>
<td>CZW1 NKL1</td>
<td></td>
</tr>
<tr>
<td>CZW3 FZZ2</td>
<td></td>
</tr>
<tr>
<td>CZW1 NAK1</td>
<td></td>
</tr>
<tr>
<td>CZW7 NAK2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: The form of a structure file (see: left), the form of a geometry file (see: right)
SUMMARY

At the present stage of a software development a SHAFT–CAD program works in the NX 6.0 environment and a CAPP system works beyond this environment. It seems to be the biggest disadvantage of this solution. Moreover implementation of the GRIP language in order to record a design features database is not the best solution because in this case is not possible to implement object paradigm fully. Additionally this solution does not support the bidirectional information exchange between CAD and CAPP systems. In order to solve this problem author of this work considers:

- based on worked out open structure of design features create a new one CAD module working in NX 6.0 environment but now implemented with OPEN/API C++ language. In order to make this module, elaboration of an object classes structure is needed, a classes structure working in a CAPP system can be used after making same modifications;

- a CAPP system should work in the NX 6.0 environment, should be written with OPEN/API C++ language.

If above assumptions will be satisfied the bidirectional information exchange between CAD/CAPP systems will be performed. It means that at the design stage a designer will have full information about design cost and he will be able to make same changes in a design during a manufacturing process planning stage. A CAPP system with its object design representation will give some information for a CAM module so full integration between CAPP system and CAM module will be performed.

ACKNOWLEDGEMENT

The presented research work was financially supported by Polish Ministry of Science and Higher Education in the framework of the N115 081 31/1906 project entitled: Integrated system of adding of a technical-organizational production preparation for small and medium enterprises.

REFERENCES


A STEP-NC Process Planning System for CNC Turning Operations

X. Zhang¹,², R. Liu², A. Nassehi¹, S. T. Newman¹

¹Department of Mechanical Engineering, University of Bath, Bath, BA2 7AY, UK
²School of Mechanical Engineering, Shandong University, Jinan, 250061, China

ABSTRACT

Over the last 50 years there have been many significant enhancements in computer aided systems which have influenced CNC technology. One area that can be considered as a bottleneck to these CNC enhancements and in particular to interoperability in CNC manufacturing is G&M part programming (ISO6983). To overcome this bottleneck, the new standard ISO 14649, known as STEP-NC, is being developed to provide detailed information on component design, process planning and machining strategies to manufacture parts for the next generation of intelligent CNCs. This standard forms the basis of a new paradigm shift in the CNC domain to support the digital modeling of CNC manufacturing resources. The research in this paper aims to identify major issues and develop new software tools to demonstrate the feasibility of interoperable CNC manufacturing based on STEP-NC. This paper proposes a Process Planning System (PPS) with surface roughness chosen as the process planning objective. There are five modules in the system: program reader, process planner, STEP-NC CAD viewer, STEP-NC CAM viewer and program writer. The reader is responsible to interpret the geometry and the manufacturing data from a STEP-NC file into a stored data list. The process planner uses this data list and enables users to evaluate surface roughness based on a mathematical model. Through the STEP-NC CAD viewer, the part geometry can be shown and via the STEP-NC CAM viewer the toolpath can be verified. Finally, the writer converts the stored STEP-NC data of the system into an updated STEP-NC file. An example case study component is used to demonstrate the PPS and show the interfacing of the STEP-NC data.

1. INTRODUCTION

From its emergence in 1952, the numerical control machine tool has undergone significant improvements and has provided an ever-increasing important part in manufacturing. Many other relating technologies, including CAD, CAM, CAPP, have advanced greatly coupling with the enhancement of the computer technology. However, the early NC machines and today’s modern CNCs utilize the same standard for programming, namely G&M codes formalized as ISO 6983. It is a common consensus that the unchanged standard has become a bottleneck of the advancement of CNC manufacturing because of the data noncompliance through the CAD/CAPP/CAM/CNC chain [1]. To eliminate the problem, a new standard extending STEP into manufacturing has been developed, which is formalized as ISO 14649 commonly known as STEP-NC. Unlike the G&M codes, which program the movements of the machine axes and switches, STEP-NC is a comprehensive data model that associates the machining objectives (CAD design data) with solutions (CAM process data required) in an object-oriented way. Furthermore, it builds up a bi-directional information highway between the CAD/CAM and CNC systems [2] and makes interoperable process planning and manufacturing possible [3]. Although it is believed that the new standard will undoubtedly introduce interoperability into manufacturing, due to the inherent complexity of STEP compliant CNC controllers, a global interoperable STEP-NC system is still a long way away [4].

Recently, there have been a number of major developments achieved in the STEP-NC based interoperable research. In the research area of STEP-NC compliant controller, a TurnSTEP has been developed by Suh et al. [5], which can convert STEP-NC part program into tool motion to get the designed part. Xu et al. [6] developed a STEPcNC converter used on an existing CNC machine tool to enable it to be G-codes free. Heusinger et al. [7] present a methodology for implementation of a CAx chain for rotational asymmetric parts. The necessary data models were created and tested through a prototype system. A framework for a multi-agent-based STEP-NC...
controller was proposed by Lan et al. [8] with a prototype system has been developed with the tool of ST-developer for the framework. In the research area of STEP compliant process planning, a three-stage process planning model was proposed by Liu et al. [9] using a feature-based approach, which divides process-related issues into three levels: offline process planning, shop-floor process planning and real-time process planning.

STEP-NC provides a real opportunity and challenge to promote the improved manufacturing capability utilizing high level and detailed information. To date, most STEP-NC systems focus on milling operation. In this research, a Process Planning System (PPS) for the next generation controller for turning is proposed. This paper introduces the PPS and outlines the mathematical model on which process planning is carried out. Finally, an example component is used to demonstrate the application of the system.

2. STEP-NC PROCESS PLANNING SYSTEM(PPS) ARCHITECTURE

For the next-generation interoperable CNC controller, a multi-agent-based intelligent STEP-NC controller has been proposed by Lan et al [8], which is comprised of many agents. These agents undertake specific tasks and accomplish them by interacting with others to carry out the various functions of an intelligent controller in a desired manner. The STEP-NC Process Planning System (PPS) which has been adapted from this approach consists of three agents based on this architecture namely: the program interpreter agent (STEP-NC data reader), process planner agent and program generator agent (STEP-NC data writer). The emphasis of this research is the implementation of the process planning agent.

PPS consists of five modules namely: program reader, process planner, STEP-NC CAD viewer, STEP-NC CAM viewer and program writer as shown in Figure 1. In the process planner module, project optimization is made by adjusting specific parameters and the workingsteps’ sequences, etc. Process planning has numerous criteria that can form the objective of the plan (e.g. minimum processing time, cost etc.). To limit the scope of this research, the surface roughness has been chosen as optimization objective. With the mathematical model outlined in section 3.2, the user compares different machining strategies with different parameters (e.g. feeds, tools insert types etc.). The intended toolpath of the machine is generated by the toolpath generator. The toolpath planned can be shown and evaluated in the STEP-NC CAM viewer, as described in section 4.

The basis of PPS is the STEP-NC data and consequently, the system contains other function modules including program reader, program writer and the STEP-NC CAD and CAM viewers. With the input of the system being a STEP-NC Part21 file, the data reader checks it and translates it into an internal data format (Java ArrayList). The data is organized by the relationship inherited by the STEP-NC file, utilizing the advantage of the object-oriented programming language Java, such as strong types, inheritance and polymorphism [10]. The process planner provides a graphical interface to display and modify the particular information associated with each feature and operation. After changing the manufacturing parameters, it is essential to check whether these changes contribute to enhance the final part based on the roughness mathematical model. Then, the updated data list could be transferred to the Data Writer to generate a new STEP-NC compliant program with the manufacturing information stored in ISO 14649 format. The ultimate output is a STEP part21 physical file (.stp file). During the process, there are several visualization tools that have been realized to view the STEP-NC data, the final part, the manufacturing features and the toolpath generated by the toolpath generator.

3. PROCESS PLANNER

3.1 STEP-NC PROCESS PLANNER

According to Liu et al. [9], process planning is divided into three levels: off-line, shopfloor and real-time. At the off-line process planning level, a NC program is generated with the CAD design information and other technology requirements of the final part. The activities involved in shop-floor process planning consist of:

i) Manufacturability Analysis: assessing the whole machining ability with all the resources available.

ii) Workingstep-level Planning: optimizing the parameters for every single workingstep.

iii) Part-level Planning: determining the best workingstep sequence for a particular project.

iv) Manufacturing Resource Planning: choosing suitable tools, clamps etc.
At the level of real-time process planning, the toolpath and additional machine function switches (e.g. turn coolant on/off etc.) are produced according to the current machine status. PPS is applicable to both the shopfloor and the real-time process planning levels. The process planner module can be applied at the shopfloor process planning level while the toolpath generator would typically be used at the real-time level.

Manufacturability Analysis is achieved by the operators according to their experience, while the other three activities are achieved by PPS through system interfaces to modify the stored STEP-NC data. These interfaces are shown in Figure 2 and 3. The first interface (Figure 2) is used to adjust the workingstep sequence and the interface in Figure 3 is for workingstep parameters changing, which can generate a surface roughness report.

In Figure 2, the interface shows for SETUP 1, with its workingsteps listed in the WS-IN and the workingsteps in the WS-REST area related to other setup(s). Using this interface, the user can add workingsteps into the chosen setup and delete workingsteps from the chosen setup. The user also can adjust the sequence of the workingsteps within a setup. If another setup is chosen, the WS-IN and WS-REST data will change correspondingly.

In Figure 3, the parameters of a workingstep are visualized in different groups. It is feasible to change some of them according to the user’s experience. Before saving the changes, the user can gain help by pressing the button of roughness report from the surface roughness prediction mathematical model to determine whether the changes contribute to enhance the quality of the part. The pop-up shows the surface roughness report based on the data of the workingstep. The surface roughness prediction mathematical model is outlined in next section.
3.2 SURFACE ROUGHNESS MODEL

There are many different options for the workingstep parameter optimization based on a wide range of goals. As described previously, surface roughness has been chosen to be the optimization objective in this research. As is commonly known, the surface roughness is mainly caused by the height of the residual area, which makes the surface when magnified look like a threaded uneven pattern. The roughness value of a part can only be obtained by measurement after it is completed, which is merely the criterion to determine whether it is qualified or not. From using the detailed workingstep and operation data in STEP-NC, it is possible to provide a standardized approach to predict and verify the surface quality of the product. As there are many elements that influence the surface quality of a part, with many of them being dependent on a particular manufacturing resource, the process planner can at most reach a level to compare different ways of machining using the mathematical model of the residual area’s height as depicted in Figure 4.

In turning operations, the tool and the workpiece have a relative helical motion which inevitably results in the residual area on the surface. The surface roughness is mainly caused by the height of the residual area together with the influence of other factors. Theoretically, the residual area height is the surface roughness. The height of the residual area depends on the major tool cutting edge angle (in STEP-NC, it is called side cutting edge angle) $K_r$, minor tool cutting edge angle (in STEP-NC, end cutting edge angle) $K'_r$, tool nose radius $r$ and the feed rate $f$. According to the relationship of the values of tool nose radius and feed rate, the residual area height is shown in three different cases as shown in Figure 4 [11]:
For a large tool nose radius with a low feed rate \((f < 2r \cdot \sin K_r')\), as shown in Figure 4 (a), the residual area height depends on the \(r\) and the \(f\), where \(r\) is the tool nose radius and \(f\) is the feed rate:

\[
h = \frac{f^2}{8r} \tag{1}
\]

For a high feed rate with a small tool nose radius \((f \geq 2r \cdot \sin K_r')\), as shown in Figure 4(b), the residual area is thought to be caused by the tool nose \(r\) and the minor cutting edge \(K_r'\). The height is calculated as follows:

\[
h = r(1 - \cos K_r' + F \cdot \cos K_r' - \sin K_r'\sqrt{2F - F^2})
\]

where \(F = f \cdot \sin K_r' / r\).

When \(r=0\), as shown in Figure 4(c), the residual area height depends on the feed rate \(f\) the major tool cutting edge angle \(K_r\) and the minor tool cutting edge angle \(K_r'\), where \(r\) is the tool nose radius and \(f\) is the feed rate:

\[
h = \frac{f}{c \tan K_r + c \tan K_r'} \tag{3}
\]

(a) for a large tool nose radius  
(b) for a high feed rate  
(c) \(r=0\)

Figure 4 The residual area height \((h)\)

In the case of the example shown in Figure 3, Equation 1 is used above with the values of \(r=1.0\, \text{mm}, K_r'=10^\circ\) and \(f=0.3\, \text{mm/rev}\). This as shown from the roughness report dialogue gives a value for \(h\) of 0.01125mm.

4. TOOLPATH GENERATOR

The STEP-NC CAD and CAM viewers provide the user with a virtual and graphical interface of the part geometry and the machining toolpath. The major capability developed in these viewers is the toolpath generation to get the tool cutters center lines.

4.1 THE SELECTION OF THE TOOLPATH PLANNING UNIT

There are three feasible options for the toolpath planning unit based on STEP-NC, which are the manufacturing-feature, machining-operation and workingstep. According to the STEP-NC standard, manufacturing features contain three different kinds of information: geometry shape dimension, quality requirement, workpiece information including material category, dimensions etc. These three forms of information do not provide enough data to generate the machining toolpath. The other decisive factors of the toolpath are the turning technology and strategy, which are stored in machining-operations. Both the manufacturing-feature and machining-operation needs to be combined with each other to generate the final path. Consequently, the workingstep, containing the manufacturing-feature and the machining-operation, has been chosen to be the path planning unit. It can be described as using an operation to machine a feature. Additionally, the workingstep is an executable unit in STEP-NC, which means, the consequence of workingsteps is the correct procedure to undertake the machining of the product.
4.2 The procedure of toolpath planning

4.2.1 Material Removal

Firstly, it is essential to determine which piece of material to remove from the stock. In the example program of ISO14649-12 as shown in the Figure 5 for the case study part shown in Figure 6, line #40 represents a workingstep in the program, with a feature (#12) and an operation (#20) defined. According to the feature of #12, its workpiece is defined in #1, and its feature outline is defined in #200.

From the program line #4 in Figure 5, it is obvious that the rough workpiece is circular with a length of 175mm and a radius of 37.5mm. The feature outline is defined by five points in its own coordinate system, which is shown in the Figure 7 [12]. It is thus easy to determine what material needs to be removed from this workingstep.

4.2.2 Machine Strategies

In STEP-NC, to identify the material removal strategy, the attribute of the operation named machining strategy is used. Examples of machining strategies defined for turning are namely (a) unidirectional turning, (b) bidirectional turning, and (c) contouring turning as shown in Figure 8. As previously discussed, the machining strategy is crucial to the toolpath planning, which is specified by the operation. Within the workingstep in the manufacturing workingstep segment (#40), the operation (#20) to rough the feature (#12) in the program is also shown in figure 5.

In this machining operation segment (#20), the first $ identifies that the toolpath of the operation is not specified yet. This is due to the fact that it is expected to be generated at the intelligent STEP-NC CNC. The retract plane is represented by the number 30.000 in millimeter, and the machining allowance of this operation is identified by the number 0.5 in millimeter.
The machining strategy is defined in line #131. From this line, the following information is specified:

i) Unidirectional turning.

ii) Multiple turning. This is omitted but could be determined by the controller intelligently with the final incremental cutting depth being 3.000mm and the height of the tool retraction height being 2.000mm. While the maximum depth to be removed is apparently much bigger than 3.000mm, consequently, the property value is true.

iii) Cutting depth list. Each entry of the cutting depth list stands for the thickness (not the total depth) that shall be removed in the respective pass until all the material to be removed is removed. If there are more passes needed than list entries, the last entry shall be used for the remaining passes [12]. In this program, the 3.000 is the cutting depth of the last pass, and the depth for other passes should be implemented by the controller. As shown in Figure 7 above, the maximum depth is 15.96mm. The whole cutting depth could be assigned as (0.96, 3, 3, 3, 3, 3) in mm.

iv) Feed direction. The direction could be specified to be parallel to the Z-axis pointing to smaller values of Z.

v) Lift direction and stepover direction. Because they are both not given in the program, in this paper, the ordinary manner is the choice: vertical lift and horizontal retreat.

With the parameters above, the final toolpath could be calculated as shown Figure 9. In the similar way, the toolpath of the rough operations in the first setup could be illustrated in the STEP-NC CAM viewer as shown in Figure 10.
6. CONCLUSIONS

This paper has proposed and demonstrated a process planning system based on STEP-NC, which can serve as the interpreter and process planner for the whole architecture of the STEP compliant NC controller. The major conclusions of this work are:

i) The STEP-NC data can be used for generic a standardized data for genetic process planning and toolpath generation for turning operations.

ii) The incorporation of the surface roughness module the STEP-NC PPS can be integrated with STEP-NC data to generate process plans which manufactures part with a prediction of surface roughness.

This research illustrates the potential that a STEP-NC data model provides the basis for future detailed standardized process planning and verified machining.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by a number of grants for Engineering and Physical Sciences Research Council (EPSRC), involving a large number of industrial collaborators. In particular, current research is being undertaken as part of the EPSRC Innovative design and Manufacturing Research Centre at the University of Bath (reference GR/R67507/01). The work is also supported by the Foundation for Outstanding Young Scientists of Shandong Province, China (2007BS05005). The authors gratefully express their thanks for the advice and support of all concerned.

REFERENCE

Mathematical model of 6-axis CNC spiral bevel and hypoid milling machine

Piotr Skawinski¹
Warsaw University of Technology
The Faculty of Automotive and Construction Machinery Engineering, Warsaw, Poland

**ABSTRACT**

*Spiral Bevel Technology Group (SBTG) is utilized in the design and manufacturing of spiral and hypoid bevel gears. The CAD/CAM system developed by SBTG for design and manufacturing of spiral and hypoid bevel gears generally supports conventional machines. This modern technology of spiral bevel gears engages the CNC milling machine or grinding machine. Therefore, it is necessary to add to the computer system a machine setup for CNC milling machines. The six independent orthogonal mechanical freedoms (X, Y, Z, A, B and C) set the relations between the cutter head and work-piece. The mathematical model of CNC milling machine accepts basic technological conventional machine settings and transforms these settings to the six axes. The model of the CNC milling machine is determined by specific vectors.*

1. **INTRODUCTION**

The base for each cutting of spiral bevel gears is generated in the machine space technological setup which contains generating gear and a blank work-piece. The generating gear is represented by cutter head and cradle. The generating gear and work-piece are rolled accordingly to the ratio of roll and the tooth profile cutting. Generally, on the conventional machine there are two types of generating gears: with flat crown gear and cone crown gear (Figure 1). CNC milling machines should allow manufacture gears in all technological methods which are accessible in conventional machines. It means that these machines should allow the ability to generate in their space the same types of generating gears.

![Figure1: Generating gear [3] for conventional machine: a) flat crown gear, b) cone crown gear](image)

¹ Corresponding author: Tel. +48 22 234 86 81, fax +48 22 234 86 22, e-mail: psk@simr.pw.edu.pl
Mathematical model of 6-axis CNC spiral bevel and hypoid milling machine

2. CONVENTIONAL MACHINE

A conventional machine for spiral and hypoid bevel gears has two units: cradle unit and work unit mounted on the machine frame. The cradle unit is equipped with a tool spindle, and depending on machine construction, the cutter axis is fixed parallel to cradle axis. The cradle unit includes the modified roll mechanism (Figure 1a) or the special mechanism that allows tilting (Figure 1b). The schematic of conventional machine with tilted cutter axis is shown in figure 2. The machines with the cutter tilt unit are very often equipped with a modified roll mechanism. It means, that such a machine allows usage of all technological methods for generating gears and formatting gears.

![Figure 2: Conventional machine [4]: 1- cradle units, 2 – cradle, 3 – eccentricity, 4 - swivel angle, 5 - cutter spindle rotation angle, 6 – work-piece unit, 7 – blank offset, 8 – frame](image)

The settings of conventional machine are following:
- cradle angle – angular position of cutter at calculation point,
- eccentricity - radial position cutter axis (set as angle of eccentricity),
- swivel angle - angle of cutter axis orientation,
- cutter spindle rotation angle (tilt) – angle between cradle axis and cutter axis,
- work offset (hypoid) – distance in vertical plane between cradle and work axis,
- root angle – the angle in horizontal plane between cradle and work axis,
- head setting – linear adjustment along work axis,
- sliding base – linear adjustment along cradle axis – the depth of cut,
- ratio of roll – decimal ratio between cradle and work (tooth profile generation),
- index ratio – only for continuously cutting (face-hobbing method) – controls the relation between number of group of cutter blades and number of teeth,
- helical motion – for technological method with helical motion during rolling motion.

The first eight settings determine relative position of the cradle and work-piece in the machine space and these settings do not change during the cutting process. The remaining settings determine the kinematics of generating
process. It is necessary during a generation process to change the following settings: cradle angle as a result of rolling motion and accordingly to this ratio of roll, pinion rotation. If it is done under continuous cutting (face hobbing), than the index ratio controls this relation between number of group of cutter blades and number of teeth in opposite to single indexing cutting. The index ratio independently of face hobbing or face milling has no influence on the generating motion (tooth profile).

3. BASIC TECHNOLOGICAL SETTINGS

Basic technological settings will be described for machines with cutter axis tilt as a general solution. In this case, a cone crown gear is generated in the machine space. For cutter tilt equaled to zero, a flat crown gear is generated in the machine space and, therefore, a swivel angle has no influence on generating gears. Figure 3 shows schematically this basic technological setup.

Figure 3: Izometric view of basic technological setup.

Figure 4: XZ plane of basic technological setup
Based on the analysis of the basic technological setup (Figures 3, 4, 5) it is necessary to define at the start some important points. Point $O_M$ is the machine center. Point $O_G$ is on the cutter axis at the plane tangent to the top of blades. Point $O_P$ is located on the pinion (gear) axis and is the top point of the pitch cone. These three points allow definition of the following settings:

- cradle angle $q$ in the XY plane (cradle plane) defined by radial setting $U$ and $X$ axis,
- radial setting $U$ as a distance in the cradle plane (XY plane) from machine center $O_M$ to cutter center $O_G$,
- swivel angle $j$ which defines the direction of cutter axis tilt. This angle is determined as a projection angle on the cradle plane and lies between perpendicular plane to the radial setting $O_MO_G$ and the plane $\lambda$ in which the cutter axis is tilted.
- cutter spindle rotation angle, tilt angle $I$ formed by cutter axis which lies on the $\lambda$ plane and cradle axis.
- root angle $\delta_M$ in the horizontal (XZ) plane formed by work axis and cradle plane XY.
- work offset $a_M$ as the distance between the cradle axis and work-piece axis. This distance is measured up or down to machine center $O_M$.
- head setting $X_P$ is the distance in the XZ plane measured along work axis between $O_P$ point and $O_{PM}$. Point $O_{PM}$ is the crossing point of work axis and parallel line to the $Z$ axis in $a_M$ distance and perpendicular to Y axis.
- sliding base $X_b$ as a distance between machine center $O_M$ and point $O_{PM}$ in perpendicular direction to the cradle plane.

These values determine the virtual machine settings as linear and angular parameters and describe the relative geometric positioning between cutter head and work-piece. The virtual machine settings should be completed in the kinematic relations between cutter and work-piece for the hobbing and milling methods, first: as the rotational position which complies the ratio of roll motion, second: as the rotational relation between number of blade groups of cutter and number of teeth for continuous indexing (hobbing method - Oerlikon/Klingelnberg technology). For the Formate methods (Gleason technology) and for the gear cutting when the cradle motion is fixed, the virtual machine settings are described only by eight geometric parameters.
4. Mathematical model of 6-axis CNC machine

Generally, the work space of CNC cutting machines is determined by the Cartesian coordinates. The number of coordinates depends on number of axes which are necessary to note the relation between tool and work-piece. For example, for a simple turning operation two axes are necessary; however, for the milling operation two, three or more (maximum five) axes are necessary. For the CNC bevel gear machine six independent coordinates describe the relative position between the cutter head and work-piece. The schematic drawing of the 6-axis CNC bevel gear machine is shown on Figure 6.

CNC machine uses 6 axis: three linear and three rotational axis which are as following:
- X - horizontal movement perpendicular to cutter head axis,
- Y – vertical movement perpendicular to cutter head axis,
- Z – horizontal movement parallel to cutter head axis,
- A – rotation of work-piece about its axis
- B – rotation of the work head about vertical axis parallel to the Y axis,
- C – rotation of the cutter head about its axis.

The mathematical model of the 6-axis CNC bevel machine is the registration of relative positioning of the tool and work-piece in which continuous changes occur between the axes. According to Figure 7 and 8, point O_M is the beginning (zero point) of the machine coordinates. These machine coordinates are the global coordinates in the machine work space. In opposite to a conventional machine, the CNC machine does not have a work-piece offset as the mechanical displacement of work-piece unit (see Figure 2). It means that the work-piece unit is more rigid than the same unit in the conventional machine. In the CNC solution, the work-piece offset is realized by the movement of the local coordinates. Typical cradle for the conventional machine is replaced with the special unit, in which elements move in perpendicular direction as X and Y axes (Figure 6). The XY plane is the cradle plane and the continuous changes on each axis effect the cradle rotation. The Z axis is perpendicular to the XY plane and parallel to the cutter axis and it controls the movements of the work-piece unit (table unit) towards the XY plane (the cradle plane). The CNC machine does not have a cutter tilt unit, but it is possible to realize the technological methods with cutter tilt as SFT (Spiral Formate Tilt), HFT (Hypoid Formate Tilt), and other similar approaches. Because of the continuous transformation of coordinates (Figure 7) by a controller during generating motion it is possible to hold the cutter axis and work axis at the same relative orientation, like in the conventional machine. The B axis of the CNC machine is the rotation about Y and it corresponds to root angle from the basic technological setup. Axes A and C require an additional explanation. For single indexing (Gleason technology), the rotational position of work-piece depends on generating motion connected with the XY cradle position and does not depend on the rotational cutter position. Continuous indexing requires, independently of generating motion (Oerlikon/Klingelnberg
Mathematical model of 6-axis CNC spiral bevel and hypoid milling machine

technology), a very close relation between angular position of the number of blade groups of cutter and number of cutting teeth. In both cases, the C axis is controlled by cutting speed. A mathematical model of CNC machine is supported on the basic technological settings for the conventional machine which have been listed above (10 settings), but determined in the 6 - axis (6 mechanical freedoms) of the Cartesian coordinates.

Figure 8: Coordinates system for 6-axis CNC machine

Vector $\mathbf{PO}$ along work-piece axis is defined:

$$\mathbf{PO} = [-\cos \delta_M, 0, -\sin \delta_M]$$  \hspace{1cm} 3.1

Vector $\mathbf{t}$ along tool axis:

$$\mathbf{t} = \left[ -\sin(i) \sin(q-j), \sin(i) \cos(q-j), \cos(i) \right]$$  \hspace{1cm} 3.2

The unit vector $\mathbf{p}$ and $\mathbf{t}$ of work-piece and cutter position:
Mathematical model of 6-axis CNC spiral bevel and hypoid milling machine

\[ \mathbf{p} = [-\sin\delta_M, 0, \cos\delta_M] \]
\[ \mathbf{r} = [-\sin(q-j), \cos(q-j), 0] \]

Vector from point \(O_G\) to point \(O_{PM}\):

\[ \mathbf{Q} = [-U \cos q, (U \sin q - a_M), x_B] - A_{MDt} \mathbf{P}\mathbf{O} + H \mathbf{r} \]

where:
- \(U\) – radial setting
- \(q\) – cradle angle
- \(i\) – tilt angle
- \(j\) – swivel angle
- \(\delta_M\) – root angle
- \(a_M\) – hypopid offset
- \(x_B\) – sliding base
- \(A_{MDt}\) – mounting distance
- \(H\) – height of cutter

Unit vectors of orthogonal axes are determined:

\[ \mathbf{\hat{W}_Y} = \frac{\mathbf{P}\mathbf{O} \times \mathbf{r}}{|\mathbf{P}\mathbf{O} \times \mathbf{r}|} \]
\[ \mathbf{\hat{W}_Z} = \mathbf{r} \]
\[ \mathbf{\hat{W}_X} = \mathbf{\hat{W}_Y} \times \mathbf{r} \]

Than, the linear coordinates \(X\), \(Y\) and \(Z\) are equal:

\[ X = -\mathbf{Q} \cdot \mathbf{\hat{W}_X} \]
\[ Y = \mathbf{Q} \cdot \mathbf{\hat{W}_Y} \]
\[ Z = \mathbf{Q} \cdot \mathbf{\hat{W}_Z} \]

The angular coordinate \(B\) means rotation about \(Y\) axis and corresponds to root angle \(\delta_M\) and is given by:

\[ B = \sin^{-1}(-\mathbf{P}\mathbf{O} \cdot \mathbf{\hat{W}_X}) \]

For the Gleason system, the \(C\) axis means the cutting velocity and the rotational angle does not depend on rotational position of the work-piece. The \(C\) axis is very important for Oerlikon/Klingelnberg system, where the relation between rotational position of cutter and work-piece exists. Thus, the reciprocal rotational position of cutter and work-piece is determined by the following functions:

- for cutter head angular position:
  \[ \omega_C = \sin^{-1}(-\mathbf{r} \cdot \mathbf{\hat{W}_Z}) \]
- for work-piece angular position:
  \[ \omega_{PO} = \sin^{-1}(-\mathbf{p} \cdot \mathbf{\hat{W}_Y}) \]

Rotation of the cutter head with number of blade groups should be synchronous with number of work-piece teeth. The rotational velocity of the cutter and work-piece does not have any relation and it is the cutting speed only.
5. CONCLUSIONS

The CNC spiral bevel gear machine that is controlled in the 6-axis has the common basic technological parameters with the conventional machine. It is very important, due to the fact that computer calculations for the conventional machines and particularly the basic technological settings are the grounds to determine relations in all 6 axes. The basic technological settings are the input data to CNC controller. Even though the cutter axis is parallel to the Z axis and perpendicular to the XY plane, the continuous transformation of the coordinates allows the possibility of using the technological methods with cutter tilt. The CNC machines allow cutting the spiral and hypoid bevel gears independently of a cutting system such as Gleason or Klingelnberg/Oerlikon.

REFERENCES


Financial support of Structural Funds in the Operational Programme - Innovative Economy (IE OP) financed from the European Regional Development Fund - Project No POIG.0101.02-00-015/08 is gratefully acknowledged.
Corporate Social Responsibility – a new trend in company management

Marcela Srchová1*, Lilia Dvořáková2, and Tereza Kadlecová3

1* Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

2 Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

3 Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

ABSTRACT

This paper presents the issue of Corporate Social Responsibility with its characteristics, assets, advantages and disadvantages. It analyses stakeholders, their importance and possibilities of their engagement as a factor of the performance and competitiveness of a company in the 21st century. This paper also seeks to answer the question: Why is CSR important and why follow requirements of CSR?

1. INTRODUCTION

The concept of CSR was established in the 1950s. It is a way of managing a company and making relationships with partners which contributes to improving the reputation and credibility of a company. The European Union supports CSR activities in the Europe. This issue has been worked on since the 1990s. The European Union uses CSR as a tool for achieving tenable and acceptable growth.

Corporate Social responsibility can be defined as a voluntary commitment on the part of a company to behave responsibly with the respect to its surrounding environment and the society within which it operates. The company conducts business in accordance with high ethical standards, builds good relationships with its business partners, cares for its employees, supports its local region, tries to minimize any negative impact on the environment from its operations.

CSR promotes the active engagement of stakeholders in the company’s decision-making processes. The term stakeholders is used to describe all parties that are involved with the company. They include people or groups of people operating both inside and outside of the company, such as owners, investors, employees, customers, business partners, the local community, etc.

2. CORPORATE SOCIAL RESPONSIBILITY

2.1. CONCEPT OF CORPORATE SOCIAL RESPONSIBILITY

CSR is based on three fundamental principles. They are economic, social and environmental. Each part of CSR contains a lot of different activities depending on the type of enterprise and the requirements of stakeholders.

---

1* Marcela Srchová: Tel.: +420 377 638 464; Fax: +420 377 638 402; E-mail: srchova@kpv.zcu.cz
The main features of CSR are:
- triple-bottom-line – economic, social and environmental,
- voluntary – all activities are done voluntarily,
- stakeholders dialogue – integration of all participants,
- long-term period – all activities are done over a long-term period,
- credibility – increasing company credibility.

### 2.2 Advantages of Corporate Social Responsibility

Responsible behaviour is advantageous, brings a lot of benefits, especially non-financial, which are also very important for successful enterprise.

<table>
<thead>
<tr>
<th>CSR advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Higher reputation of company, better company image</td>
</tr>
<tr>
<td>• Higher attraction for investors</td>
</tr>
<tr>
<td>• Good reputation and strong position in market</td>
</tr>
<tr>
<td>• Distinguish from rivals</td>
</tr>
<tr>
<td>• Increasing employees’ productivity and loyalty</td>
</tr>
<tr>
<td>• Attraction for quality and talented potential employees</td>
</tr>
<tr>
<td>• Decreasing expenses on risk management</td>
</tr>
<tr>
<td>• Direct financial saving – ecological behaviour</td>
</tr>
<tr>
<td>• Better relationships with local society and public institutions</td>
</tr>
</tbody>
</table>

CSR does not bring advantages only to big companies, small and medium sized companies can also gain some advantages. They are: easier recruitment of employees and business partners, good public relations, possibility of getting contacts and information from society and from main business partners.

### 3. Stakeholders

#### 3.1. Classification of Stakeholders

A company has three kinds of stakeholders: organizational stakeholders (internal to the company) and economic and societal stakeholders (external to the company).

*Organizational Stakeholders*
First, stakeholders exist within the organization. The business functions and practices that affect organizational stakeholders are presented from a CSR perspective. Examples of organizational stakeholders are stockholders, employees, and managers.

**Economic Stakeholders**

Second, they serve as the interface between the organization and its larger social environment. Not only do the issues in this section affect the financial aspects of the organization, they create bonds of accountability the firm and its wider operating context. Examples of economic stakeholders include consumers and creditors.

**Societal Stakeholders**

Third, topics that affect societal stakeholders determine the business environment in which the firm operates. Examples of societal stakeholders include government agencies and regulators, communities, and the environment itself.

In this figure you can see the stakeholder classification.

![Figure 2: Stakeholder Classification](image)

This model is not limited to for-profit organizations and can be applied equally to governments and nonprofits. Economic stakeholders serve as the accountability interface between any organization and the society within which it operates. Among businesses, the accountability mechanism is profit and loss. If a company fails to make a profit over the long term, it will be unable to remain in business. For democratic governments, the accountability mechanism is twofold. The campaign contributions that finance the democratic process and the votes of stakeholders (taxpayers), which determinate the outcome. And for nonprofits, the accountability mechanism is the organization’s ability to generate sufficient operational funds by justifying its continued existence to economic stakeholders.

Without these economic interfaces, organizations lose their mechanism of accountability and, therefore, their legitimacy over the long term. This is true regardless of whether the organization is a business, government, or nonprofit.

### 3.2 Importance of Stakeholders

The systematic and regularized involvement of key stakeholders gives the business numerous advantages, [4]:

**Innovative Environment:** A continuing dialog with stakeholders represents a valuable tool for the prediction of new trends and possible areas of interest to the business. It allows the company to better understand the complexity of the business environment and current market trends, helping it to properly put together its future strategic plans.
Risk Management: A dialog with external parties can often warn a business about potential future risks – this is especially true if the company is able to maintain a relationship with those stakeholders having a negative perception of the company.

Mutually Beneficial Relationships: Face-to-face contact and the ability to maintain a close relationship with stakeholders is the best way to build mutual trust between the company and its key stakeholders.

Setting of Objectives and Performance Monitoring: An outsider’s perspective allows the company to properly set out its objectives and evaluate its actual performance.

Access to Information: Strong and close relationships with its stakeholders can be a valuable source of information for the company.

3.3 Engagement of Stakeholders

An explanation and overview of the roles played by the stakeholders throughout the four stages of the CSR implementation process (planning, execution, evaluation and improvement) helps to better understand how important it is to get the stakeholders involved in the decision-making processes of one’s business, [5]. The roles are briefly outlined here and showed at the following figure:

Planning
During the planning stage, stakeholders can help us to understand the respective economic, social and environmental impacts of our company’s actions and help us choose the right CSR strategy for our business.

Execution
Stakeholders play an important role in our business’ execution of our CSR strategy by helping us to define and act upon our company’s commitments in the area of corporate social responsibility.

Evaluation
Stakeholders are an essential part of the monitoring and reporting of our business' performance. They are included in our selection of performance indicators and provide us with feedback on our performance in the different CSR areas.

Improvement
Our stakeholders contribute to the shaping of ideas for improving our existing CSR strategy.

![Figure 3: Engagement of stakeholders](image-url)
4. IMPLEMENTATION OF CSR

Many companies are already engaging in activities in the area of customer and employee care, community support and environmental protection. These activities can serve as a good starting point for the development of a more strategic approach to Corporate Social Responsibility.

When implementing the CSR concept, it is important to have a systematic approach and to take the following into consideration with the respect to your company its:

- mission statement,
- corporate culture,
- area of specialization and area of business,
- business strategy,
- environmental profile,
- risk profile,
- operating conditions.

The following implementation diagram is based on the well-known PDCA concept:

- Plan = plan ahead.
- Do = execute your plan.
- Check = evaluate the results.
- Act = act on the results by trying to improve.

![PDCA concept diagram](image)

5. CSR COMMUNICATION

There are many communication tools and methods that the company can use when providing communications about its socially-responsible behavior to others. Some of these are highly affordable and take a minimum of effort, [6].
Table 2 Tools of communication CSR

<table>
<thead>
<tr>
<th>Tools of CSR communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>• CSR report</td>
</tr>
<tr>
<td>• Internet / Intranet</td>
</tr>
<tr>
<td>• Internal / external newsletter and magazines</td>
</tr>
<tr>
<td>• Recruitment process / employee training</td>
</tr>
<tr>
<td>• Corporate information boards</td>
</tr>
<tr>
<td>• Open house events</td>
</tr>
<tr>
<td>• Code of Ethics</td>
</tr>
<tr>
<td>• Standards and norms</td>
</tr>
<tr>
<td>• Product designations (employer of the year, responsible enterprise)</td>
</tr>
<tr>
<td>• Awards and recognition</td>
</tr>
<tr>
<td>• Active communication with the media (press releases, conferences, announcements)</td>
</tr>
</tbody>
</table>

Once the company has initiated a collaborative relationship with its stakeholders and started to engage in activities that are part of its overall CSR strategy, it is time to let the surrounding community and also the company’s own internal community know about its initiatives in the area of corporate social responsibility. As a first step, the company has to make sure that its employees and partners are supportive of the company’s efforts in this area. In order to make this happen, the company has to thoroughly explain to them what, how and why certain types of CSR activities are being undertaken. The company’s external stakeholders also need to be properly informed about these CSR efforts because this is the only way in which the company will be able to set itself apart from its competitors in the eyes of these external parties. Let the others know that you are a socially-responsible business. This information should be appropriately communicated, reflecting the activities and initiatives in which your company is involved.

Regularly communicating information about the company’s socially-responsible behavior can provide the company with a number of advantages. They are briefly outlined in the following table.

Table 3 Advantages of CSR communications

<table>
<thead>
<tr>
<th>Advantages of CSR communications</th>
<th></th>
</tr>
</thead>
</table>
| Transparency                       | • Provide others better information about the company’s CSR activities.  
                                      | • Make the company and its business more transparent. |
| Better control over CSR activities | • Compare proclaimed CSR commitments with what is actually taking place.  
                                      | • Identify weaknesses in its corporate CSR strategy. |
| Involvement of stakeholders        | • Strengthen the position.  
                                      | • Involvement of stakeholders in company decision-making processes |
| Non-commercial partnerships        | • Promote non-commercial partnerships between for-profit, governmental and nonprofit organizations. |
6 SUMMARY

There are a lot of argument for Corporate Social Responsibility. CSR broadly represents the relationship between a company and the principles expected by the wider society within which it operates. It assumes businesses recognize that for-profit entities do not exist in a vacuum and that a large part of their success comes as much from actions that are congruent with societal values as from factors internal to the company.

CSR is a rational argument for businesses seeking to maximize their performance by minimizing restrictions on operations. In today’s globalizing world, where individuals and activist organizations feel empowered to enact change, CSR represents a means of anticipating and reflecting societal concerns to minimize operational and financial limitations on business.

Summing the moral and rational arguments for CSR leads to an economic argument. To incorporate CSR into operations offers a potential point of differentiation and competitive market advantage upon which future success can be built, besides avoiding moral, legal, and other sanctions.

CSR influences all aspects of a business’s day-to-day operations. Everything an organization does interacts with one or more of its stakeholders groups. Companies today need to build a watertight image with respect to all stakeholders. CSR is an argument of economic self-interest for business. CSR adds value because it allows companies to reflect the needs and concerns of their various stakeholder groups. By doing so, a company is more likely to retain its societal legitimacy and maximize its financial viability, over the long term.

REFERENCES


Web-DPP: A Distributed Process Planning System for Adaptive Machining Operations

Lihui Wang1*, Ji Ma2 and Hsi-Yung Feng3

1 Centre for Intelligent Automation
University of Skövde
PO Box 408, 541 28 Skövde, Sweden

2 Department of Mechanical and Materials Engineering
The University of Western Ontario
London, Ontario N6A 5B9, Canada

3 Department of Mechanical Engineering
The University of British Columbia
Vancouver, British Columbia V6T 1Z4, Canada

ABSTRACT

Outsourcing, joint ventures, and cross-border collaborations have led to work environments geographically distributed across organizational and national boundaries. Targeting the distributed environment, the Web has been widely used for development of collaborative applications supporting dispersed working groups and organizations, because of its platform, network and operating system transparency, and its easy-to-use browser interface. The objective of this research is to develop an innovative Web-based distributed process planning system (Web-DPP) for machining operations. Our approach tries to engage a dispersed working group in a collaborative environment, allowing the team members to share real-time information through the Web-DPP. This paper presents both the system design specification and the latest development of this system.

1. INTRODUCTION

Today’s manufacturing systems are required to be flexible and adaptive to unpredictable changes in a dynamic environment, where manufacturing operations may be rearranged or replaced upon the changes. Such manufacturing systems must contain collaborative and intelligent entities that can adaptively adjust themselves so as to achieve a global system objective. As a constituent component of the manufacturing systems, the entity for process planning is required to be responsive and adaptive to the rapid adjustment of production capacity and functionality. Unfortunately, traditional process planning methods are time-consuming and error-prone, if applied to such a changing environment. Therefore, adaptive and intelligent process planning has been a hot topic for the last decade. A number of approaches have been reported in the literature, trying to solve the challenging problem in process planning. Recent efforts include object-oriented approach [1], Petri net-based approach [2], neural network-based approach [3], knowledge-based approach [4], genetic algorithm-based approach [5], and feature-driven approach [6]. These reported approaches and their combinations have been applied to several specific problem domains, such as setup planning [7], process sequencing [8], tool selection [9], cutting parameter selection [10], and tool path planning [11].

Despite the achievements in solving complex decision issues in process planning, these reported approaches and systems are limited to static problems through decisions made in advance. Their adaptability to unpredictable changes on shop floors remains insufficient. Most process planning systems available today are centralised in architecture, and off-line in data processing. It is difficult for a centralised off-line system to make adaptive decisions, in advance, without knowing actual run-time conditions on the shop floors. To be responsive to sudden changes, a distributed and adaptive approach is considered suitable for handling the dynamic situation, especially for job shop operations. In this paper, we propose a Web-based distributed process planning (Web-DPP) approach to solving the problems.

* Corresponding author: Tel.: +46 (0500) 44-8519; Fax: +46 (0500) 44-8598; E-mail: lihui.wang@his.se
The objective of the Web-DPP is to improve the performance when planning machining operations on shop floors. It uses a two-layer structure – *supervisory planning* (SP) and *operation planning* (OP). The supervisory planning focuses on high-level machining sequence generation, while the operation planning focuses on machine-specific working step planning and execution. More details of the Web-DPP concept are introduced in Section 2, followed by the architecture design in Section 3. Section 4 presents in detail the system analysis of the Web-DPP in IDEF0, which is implemented in Section 5. Finally, in Section 6, our research contributions and future work are summarised.

### 2. WEB-DPP CONCEPT

Figure 1 illustrates our long-term research activities of distributed process planning, dynamic scheduling, real-time monitoring and remote control in a shared cyber workspace, where Web-DPP is the focal point of adaptive decision making based on monitoring information and available resources from scheduling.

Process planning is the task that transforms design information into manufacturing processes and determines optimal sequence for machining. A process plan generally consists of two parts: *generic* data (machining method, machining sequence, and machining strategy) and *machine-specific* data (tool data, cutting conditions, and tool paths). A two-layer hierarchy is, therefore, considered suitable to separate the generic data from those machine-specific ones in Web-DPP. This concept is shown in Figure 2.
Other than SP and OP, machining features and function blocks are two crucial concepts adopted in the Web-DPP. They carry the machining process information and go through a number of functional modules of the system. As shown in Figure 2, machining features are first created and maintained as part of product data by a machining feature-based design system (for non-feature-based design systems, a third party utility tool is needed for features recognition). The tasks of Web-DPP are divided into two groups and accomplished at two different levels: shop-level supervisory planning and controller-level operation planning. The former focuses on product data analysis, machining feature decomposition, setup planning, machining process sequencing, jig/fixture selection, and machine selection. The latter considers the detailed working steps for each machining operations, including cutting tool selection, cutting parameters assignment, tool path planning, and control code generation. Between supervisory planning and operation planning, scheduling functions can be integrated by means of function blocks. Because of the two-level structure, the decision-making in Web-DPP becomes distributed in terms of (1) timing (supervisory planning in advance vs. operation planning at runtime) and (2) location (supervisory planning in one office workstation vs. operation planning by many machine controllers). The separation of decisions also makes the high-level process plans generic and portable to alternative machines. In other words, since a final process plan is generated adaptively yet at runtime by CNC controllers, there is no need to generate redundant alternate process plans.

### 3. WEB-DPP SYSTEM ARCHITECTURE

The Web-DPP is not limited to process planning. It also covers process plan dispatching and execution monitoring at shop floor and machine levels. Such functionalities are designed into the system architecture as shown in Figure 3, where supervisory planning, execution control and operation planning are the three major components. The execution control looks after the dispatching of jobs while interfacing with scheduling and monitoring modules to retrieve the latest information of availability of resources and status of machines. In this research, we neglect the feature-based design and feature recognition at product design stage, based on an assumption that machining features are already available in product data. They are either created directly by using a feature-based design system or recognised by a third party feature recognition solution.

- **Part Geometry or Features**
- **Product Data**
- **Supervisory Planning**
  - **Function Block Design**
- **Function Block Processing**
- **Execution Control** (Adaptive Setup Merging/Dispatching)
- **Operation Planning**
  - **Generic Setup Planning**
  - **Adaptive Setup Merging**
- **Open CNC Controller**
- **Factory Shop Floor**
- **Design Office**
- **Tool Database**
- **Resource Database**
- **Manufacturing Knowledge Base**
- **Machining Features**
- **Machining Techno.**
- **Cutting Strategy**
- **Tool Path Generation**
- **Tool Database**
- **Gateway**
- **Fieldbus**
- **Pocket Roughing**
- **ECC**
- **Corporate Network**

**Figure 3: System architecture of Web-DPP**

During supervisory planning, a generic setup plan can also be created by grouping machining features according to their tool access directions. The generic setup plan is for 3-axis machines. Necessary setup merging for 4- or 5-axis machines is conducted during the execution control and before setup dispatching to best utilise the capability of the machines. Decision making at different stages is supported by networked knowledge bases and databases, whereas the latest monitoring information is made available to the execution control for directing dynamic scheduling and dispatching machining jobs.
4. System Analysis

The system analysis work of the Web-DPP is carried out using IDEF0. The three core components of the Web-DPP shown in Figure 3 are modelled in IDEF0 in Figure 4, together with their inter-relationship and data flow, where M1 to M5 represent human, computer, network, security and machine, respectively. Meta function blocks (MFBs) are used in this research to encapsulate machining sequences (of setups and machining features), and are the output of supervisory planning. As its name suggested, an MFB only contains generic information about process planning of a product. It is a high-level process template, with suggested cutting tool types and tool path patterns, for subsequent manufacturing tasks. Execution function blocks (EFBs) are the function blocks that are ready to be downloaded to a specific machine. Basically, an EFB can be created by instantiating a series of MFBs associated with a task. Each manufacturing task corresponds to its own set of EFBs, so that the monitoring functions can be conducted for each task unit. The structure of an operation function block (OFB) is as same as an EFB. However, OFB specifies and completes EFB with more detailed, machine-specific data about machining processes and operation sequences. Moreover, operation planning module can override and update the actual values of variables in the EFB, so as to make it locally optimised and adaptable to various events happened during machining operations. We use two different terms of EFB and OFB to distinguish a given function block, because they are two separate entities with different level of detail in contents, fulfilling different level of execution, residing in different systems, and moreover, they may be deployed in physically distributed CNC controllers.

4.1. Supervisory Planning

An incoming main manufacturing plan (from a high-level production planning system) triggers the SP to generate a machining sequence plan of a given product. Feature-based reasoning is applied in the SP with considerations of only regular manufacturing resources (3-axis milling machines, regular fixtures and cutters, etc.) and the support of general knowledge base of machining technology and constraints. The generated machining sequence plan is passed to a function block designer and packed as networked meta function blocks. Details of the internal structure and data flow of SP are depicted in Figure 5. Within the SP, the function block designer is responsible of defining new function block types, specifying task-specific algorithms for each function block type, and mapping machining features to MFBs according to the generated machining sequences.

4.2. Execution Control

Scheduling information and monitoring events are integrated into and handled by the execution control module, which makes execution control an important integration point of data, activities and decision makings of the Web-DPP.
system. The functionalities of execution control include setup merging (for 4- and 5-axis machines), event handling, function block dispatching, and process monitoring as shown in Figure 6.

4.3. OPERATION PLANNING

OP is a real-time execution environment of operation function blocks. It not only specifies and optimises the process plans received from the SP (cutting tool selection, machining sequence optimisation, machining parameters selection, and tool path generation), but also executes the OFBs one by one, dynamically, in an explanation engine (Executor). In this way, the operation planning process on a machine controller can be truly adaptive, which means it can dynamically modify its process plans according to the dynamics of the actual machining process. Details of the OP are given in Figure 7.
5. SYSTEM IMPLEMENTATION

The Web-DPP is designed to use the popular browser-server architecture and VCM (view-control-model) design pattern with built-in secure session control. The proposed solutions for meeting both the user requirements of rich visual data sharing and real-time constraints are listed below.

- Based on Web B/S architecture
- Using interactive scene graph-based Java 3D models based on XML-style file format
- Providing users with browser-based thin-client graphical user interface for distributed process planning
- Deploying major planning and control logics in a secure application server

![Figure 7: IDEF0 model of operation planning](image)

![Figure 8: Three-tier architecture for Web-DPP implementation](image)
The mid-tier application server handles security concerns and control logics, including session control, viewer registration, data collection/distribution, and algorithm invocation. A central session manager is designed to look after the issues of user authentication, session control, session synchronisation, and sensitive data logging. All initial transactions need to pass through the session manager for access authorisation. A registrar is designed to maintain a list of subscribers with the requested data access. The http streaming protocol is chosen for the best combination between applets and servlets used for system implementation. Another server-side component called data accessor is designed to separate the logical and physical views of data. It encapsulates JDBC (Java Database Connectivity) and SQL codes and provides standard methods for accessing data (Java 3D models, knowledge base of the devices, or XML-style documents). Figure 9 depicts the system implementation package diagram, and Figure 10 shows one snapshot of the system for web-based distributed process planning.

![Figure 9: Implementation package diagram of Web-DPP](image)

![Figure 10: Web-DPP for distributed process planning](image)
6. CONCLUSIONS

This paper presents a Web-based approach for distributed process planning (called Web-DPP) in a dynamic and uncertain manufacturing environment, particularly for job shop machining operations. The advantage of such a system is the adaptive decision making to unpredictable changes. The system is designed to separate generic process planning information from machine-specific ones. The novelty and contributions of this research are summarised below.

• Two-layer distributed system architecture for distributed decision making,
• Machining feature based reasoning for machining sequence determination,
• Function blocks to encapsulate machining sequence for controller-level operation planning, and
• Execution control linking process plans with monitoring and scheduling for job dispatching.

Our future work will focus on algorithm development, functionality enhancement and testing through real-world case studies. Integration with a third-party scheduling system and a monitoring system is also under development, the results of which will be reported separately.

ACKNOWLEDGEMENT

The Web-DPP research is supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Implementation of Automobile Spare Tire Inserting System using Augmented Reality

Hong-Seok Park *, Jin-Woo Park, and Hung-Won Choi
School of Mechanical and Automotive Engineering
University of Ulsan
Ulsan metropolitan City, 680749, South Korea

ABSTRACT

Nowadays, the global market requires shorter life cycles and variety of products to satisfy customer’s demands. Many manufacturing companies need to design and implement manufacturing systems rapidly to enhance global competitiveness. For these reasons, VR (Virtual Reality) technology has been used principally for system design and process development of many companies in the manufacturing industry over the last few years. However, VR based digital manufacturing methods and software require perfect 3D models of whole manufacturing system in the virtual environment to implement and simulate target systems. This means that modelling tasks for new system configuration increase useless expenses and efforts. AR (Augmented Reality) technology as a new man-machine interface has been introduced by not only computer science companies but also some manufacturing companies. AR technology superimposes required virtual objects on the real-time image of the physical manufacturing environment. Therefore, the modelling time and cost for new manufacturing system can be dramatically reduced by using AR technology.

In this paper, the architecture of the developed AR software for the manufacturing system is introduced. Moreover, the methods such as a multi-marker coordinate system, a threshold image, a clipping plane for virtual objects, and so on, are proposed to generate the operation program for the manufacturing system. Base on these methods, the test bed of the automobile spare tire inserting system is configured as a 5X-reduced scale and the operation program for spare tire inserting is generated.

1. INTRODUCTION

In recent years, the computer graphic technology has been growing dramatically with computer industry growth. In this exigent situation, the research of VR (Virtual Reality) and AR (Augmented Reality) technology which describe realistic scenes is performing actively [1-5]. Moreover, not only computer science companies but also many manufacturing companies have been studying about VR technology, because the global market requires a variety of product and shorter life cycle to fulfill the diverse demands of customers to survive in the turbulent and competitive market.

In the case of VR based digital manufacturing technologies, it can be used to analyze static and dynamic system behavior at all stages of manufacturing system configuration as an advantage [6-7]. To put it the other way around, the modeling work requires much expenses and efforts, because the whole system has to be modeled. Some manufacturing companies which realize the weaknesses of VR based digital manufacturing technologies have been studying AR technology. AR technology can remarkably reduce the modeling work, because it uses the real environment to design and plan manufacturing systems [8-9].

The ambitious goal of AR is to create the sensation that virtual objects are present in the real world. To achieve the effect, software combines VR elements with the real world. Obviously, AR is most effective when virtual elements are added in real time. Because of this, AR commonly involves augmenting 2D or 3D objects to a real-time digital video image. The simplest example of visual AR is overlaying a 2D image on digital video. However, it is also possible to

* Corresponding author: Tel.: (82) 52-2592294; Fax: (82) 52-2591680; E-mail: phosk@ulsan.ac.kr
add 3D objects. Generally speaking, adding a 3D object to real-time video makes for a more impressive demonstration of AR technology [2].

AR technology is consisted basically of three modules which are image processing, tracking and rendering. The image processing module has to support the various types of the image data and can process the lots of image data. The image processing module sends the obtained image data to the tracking module. And the rendering module performs image superimposition. Among these three modules, tracking and registration problem is one of the most fundamental challenges in AR research today. The precise, fast, and robust tracking of the observer, as well as the real and virtual objects within the environment, is critical for convincing AR applications. Because of this reason, much research effort is spent to improve performance, precision, robustness, and affordability of tracking systems. Besides tracking, real-time rendering is also basic element for augmented reality. Since AR mainly concentrates on superimposing the real environment with graphical elements, fast and realistic rendering methods play an important role [3].

In this paper, Chapter 2 describes the design of AR system which can be used rapidly and easily at the process of manufacturing system configuration. The scope of the development for AR system, the architecture of AR system, the configuration of the user interface and the functions of the user interface etc. are introduced. Chapter 3 introduces the methods of software implementation based on the basic design. The basic structure of image processing and the tools for implementation of image processing module are introduced. And the information which is needed to track the coordinate data, moreover the methods for coding by using C++ is described for tracking and rendering techniques. Chapter 4 describes the usability of developed AR system to apply to the automobile spare tire inserting system.

2. DESIGN OF AUGMENTED REALITY SOFTWARE FOR MANUFACTURING SYSTEM

2.1. THE SCOPE OF THE DEVELOPMENT

In this research, the software development activities of AR for the manufacturing system consist of requirements analysis, specification, software design, implementation, testing and maintenance. These processes were performed recursively with correction of each step. In the requirements analysis step, the general requirements were gained from the client and the analysis of the scope of the development was determined clearly. The analysis document of requirements includes the flow chart of overall tasks, the systematical analysis information of functions, activities and data. The concise software development scope is as follow.

1) The user selects the camera devices which are connected to computer by the device enumeration dialog box.
2) The user inputs the each number and total count of markers.
3) The 3D objects (.wrl) are loaded to use for the parts of the manufacturing system.
4) After markers and virtual objects setting, the display device shows superimposition of real-time images.
5) The translation, rotation and scale data of each virtual object can be controlled by the user.
6) The clipping plane for collision detection can be created and also can be controlled by the user.
7) The visible and invisible functions for the images and virtual objects are able to operate.
8) The marker tracking start and stop functions can be used during the software activity.
9) The user can confirm the information about program activation by the message boxes.
10) The information of each marker and each virtual object data can be removed by the user.

Based on the specification document of requirements analysis and the three basic modules which were introduced at chapter one, the auxiliary functions can be supplemented for manufacturing environments. In this paper, the convenient functions which are clipping planes for collision detection and threshold control panel for the dim condition. And these functions support the user to work efficiently. Figure 1 illustrates the architecture of AR system. And the core development tool is MFC (Microsoft Function Class) based on C++.

The image data from a camera device is processed in video interface for the marker tracking. The video interface converts the video image stream into non-calibrated RGB24 image. Then the marker tracking function calculates continuously coordinate data of the markers to detect location of each marker. The basic coordinate system for positioning virtual objects is established through the matching procedure between the information of the marker database and the input data of the user interface. The rendering function performs generating and removing of virtual
objects with calculated coordinate data. Also, the coordinate transform of virtual objects is executed by using three translations data, three rotations data and three scales data.

Figure 1: Architecture of AR System

2.2. THE CONFIGURATION OF THE USER INTERFACE

The user interface designed detailedly based on basic functional analysis of whole system as shown in Figure 2.

Figure 2: Layout of user interface

1) The menu bar
2) The display plane for the render scene
3) The camera device controlling panel
4) The marker information setting panel
5) The virtual object controlling panel
6) The tracking function panel (start/stop)
7) The message enumerating panel to check program state
The menu bar consists of the functions which are performed from the right side panels. The display plane is the drawing board for superimposition of scenes and it draws 3D objects and 2D images from the camera. The user can confirm the list of connected camera devices by using the camera control panel and can change the camera properties easily. The marker information setting panel is used to input the information of the marker such as the size of the marker, the number of the marker, the total count of the markers to use and so on. And it is also used to control the marker information. The virtual object controlling panel is used to load the 3D objects for rendering and to control. The tracking function panel is used to control start and stop for the position tracking function. The message enumerating panel is helpful for user’s activities to check program state.

3. IMPLEMENTATION OF AUGMENTED REALITY SYSTEM

3.1. OBTAINMENT AND CONVERSION OF REAL-TIME IMAGE DATA

Precision and accuracy of target tracking depend on the image resolution of the camera device. USB Web Camera which has low resolution is used generally for experiments at a large number of laboratories. However, ordinary manufacturing system requires the tasks with high precision and accuracy. Because of this reason, the digital video camcorder that is connected to computer through the IEEE 1394 interface was used. OpenCV is the prevalent open source vision library suitable for computer vision applications [10]. However OpenCV does not support digital video image type. To solve this problem, high resolution buffer is obtained by using Microsoft DirectShow technology in this research. The rapid image processing is possible by using various filters. It is a representative advantage of DirectShow [11]. Also well-composed filters are figured out readily through the GraphEdit which is the utility of DirectShow. In this paper, device filter, preview filter, digital video supporting filter, SampleGrabber filter and Video Renderer filter are used to gather RGB24 image data. Moreover the obtained image buffer is converted into IplImage structure of OpenCV. This technique has the ability to access readily and offers the efficient method which converts RGB24 image data into different image type. The last image type is BGR24. Figure 3 depicts whole image processing step structure.

![Figure 3: Image obtaining and conversion steps](image)

3.2. POSITION TRACKING OF TARGET MARKERS

The position tracking of the target marker is the core technology that is determinant for precision and accuracy of AR technology. For tracking target markers, many methods, such as mechanic, magnetic, and optical are examined. Here, the optical method is widespread due to its high precision among the tracking methods. The camera device reads real-time video streams to generate see-through effects on the display equipment. Then edge detection is performed by thresholding with constant value. After this step, the information of detected marker is used to track the local coordinate in the centre of the marker by using matrix calculation.

There are some pose estimation algorithms as the robust planar pose algorithm, the fast pose estimation algorithm, and so on. These algorithms were ported to C++ added to the open source libraries as AR Toolkit, ARToolkitPlus, ARTag, and so on [1-3], [12]. In this paper, ARToolkitPlus tracking function that includes improved pose estimation...
Implementation of Automobile Spare Tire Inserting System using Augmented Reality

quality with less jitter and improved robustness was used to track coordinate system. The important function for using is rppGetTransMat() [1], [13]. This function performs matrix parameters translating to transmit the coordinate data to the rendering module. For this achievement, codes of the matrix parameters translating function and of the tracking function were generated.

The virtual object can be superimposition on the real scene by using each ID-based single marker. Moreover the multi-marker which has one local coordinate system can be used to track. Every single marker size was set 100mm and the user can change the marker size as required at the user interface. The multi-marker size can be changed in the designated data file.

3.3. IMPLEMENTATION OF VIRTUAL OBJECTS LOADING AND SUPERIMPOSITION OF SCENES

The superimposition of scenes is performed by using the obtaining image data, the tracking positions and the virtual objects. The system requirement of the hardware is prime concern to draw the scenes into viewing rectangle of the user interface. Therefore, the overall specification of the hardware as a graphic device, CPU and memory devices have to be high. The high specification of the hardware prevents jitter and lag of the superimposition scenes [14]. Also the VRML (Virtual Reality Modeling Language) object files decrease the software overhead and improve the problems of jitter and lag status. All things considered, the object control panel can load text files (.txt) which include the coordinate information of the virtual object based on rearrangement coordinate matrices. And the data structure including the two-dimensional array was designed for each object rendering on the marker. The data structure has the information of each marker number, visible/invisible, object number and translation-rotation-scale of each object as shown in Figure 4. These processes are performed by using OpenGL open source graphic library [15].

![Figure 4: Data structure including two-dimensional array and the implementation codes of the data structure](image)

This information can be controlled through the object control panel. The translation unit value of the virtual object is 1mm and the rotation unit value of the virtual object is 1°. In the case of scaling, required input value through the user interface is 100 when the side length of the square marker is 100mm. If the side length of the square marker is 40mm, the input value must be 40 to satisfy 1:1 scale between the viewing object and the modeling object.

3.4. IMPLEMENTATION OF AUXILIARY TOOLS

The clipping planes can be generated to check the collision between the virtual objects and to ensure the inside area of the virtual objects. The gClipPlane() function of OpenGL was used to generate clipping planes. Therefore, the user can generate 6 planes and can use them at the same time [15]. The normal vector of the clipping plane is derived from the equation of plane to achieve same effect as object controlling. The equation of plane is derived by using one known normal vector and one known point in the three axes generally. However, the normal vector of the clipping plane is derived by using one known point O(0, 0, 0) and three angles which were inputted by the user as shown in Figure 5 and Equation (1). And the point O(0, 0, 0) is always used to calculate the equation as the known point.
\[
\begin{align*}
a &= \cos(\theta_y)\cos(\theta_z) \\
b &= \cos(\theta_x)\cos(\theta_z) \\
c &= \cos(\theta_x)\cos(\theta_y)
\end{align*}
\]

(1)

where, \(\theta_x\), \(\theta_y\), \(\theta_z\) = three known angles

Through the derived normal vector decides the direction of the plane and the plane can be rotated between -180° and 180° by the user’s activity. In this case, both the translation and scaling were not considered.

The generated clipping plane does not have visualization feature. Because of this problem, there is the difficulty to recognize the location of the clipping plane in the virtual space. To solve this problem, the virtual object of the grid plane is created basically and it is matched with the clipping plane at the same time as shown in Figure 6.

The manufacturing field has poor light condition. For this reason, the optical marker cannot be recognized easily. To supplement this handicap, threshold value is controlled through the panel. The function of controlling threshold value returns selected pixel value of the binary image from 0 to 255. To control the threshold value, the functions of OpenCV are added as a cvThreshold [10].

4. Reconfiguration of Manufacturing System based on Augmented Reality

4.1. Component Modeling Work for System Configuration

The developed AR system is applied to the reconfiguration process of the automobile cockpit module assembly system which is the test bed for the real manufacturing system [9]. The target system is the automobile spare tire inserting system. Almost new components have to be remodeled. However, the air finger, the connecting part of the
robot and the ball screw are changed partial dimension except overall outer shape. In this case, the existing facility of the test bed is used as it is. The gripper and air finger which perform assembly work, the automobile body for transfer, and the jig/fixture for the automobile body, the encoder and coupling for the synchronous velocity are continuous parts of the virtual object structurally. And these parts perform similar work. Therefore, the each 3D model of them is generated. And the spare tire, the mount for the robot and the jig/fixture of the spare tire are modeled individually. Moreover, the auxiliary tools for collision-free between the peripheral unit and for accurate assembly work were modeled. The centre datum line was generated individually to match between the assembly part of the automobile body and the spare tire. And the approach path was set and was modeled to avoid collision between the robot and the automobile trunk. The operator is able to perform generating of the robot operation program precisely by using these auxiliary tools. Figure 8 shows the 3D models for system configuration.

**4.2. Setting Coordinate Systems based on Markers**

To perform superimposition of the generated 3D models, the coordinate system for positioning virtual objects has to be established. The Generated 3D models are divided into two groups. The first group has dynamic motion and the second group does not have movement. The spare tire, the gripper and air finger which perform assembly work require independent one coordinate system (coordinate system No.1). Also the centre datum line of the spare tire for assembling serviceability used the coordinate system No.1. The coordinate system No.1 has to be mounted on the end of the robot gripper. Moreover, the system should be seen in every direction due to free movement of the robot. Therefore, the coordinate system No.1 which has several markers was manufactured in the structure of a box type. And the automobile body and its jig/fixture on the carriage part require independent one coordinate system (coordinate system No.2), because the carriage plate influences to the automobile body and its jig/fixture. The centre datum line of the automobile body and the path of assembly work used the coordinate system No.2. In addition, one marker which considers the position of the camera device was established. The encoder, the coupling, the mount for the robot and the jig/fixture of the spare tire are fixed components and they have collision-free during assembly work. So these components use one coordinate system (coordinate system No.3). Figure 9 illustrates the generated coordinate systems for configuration of the spare tire inserting system.

Figure 8: Generated 3D models for spare tire insert system

Figure 9: Generated 3D models for spare tire insert system
4.3. GENERATING OPERATION PROGRAM FOR SPARE TIRE INSERTING SYSTEM

The test bed of the automobile spare tire inserting system is configured as a 5X-reduced scale. The operator generates the robot operation program with two cameras. One camera is set at the side of the test bed and the other is set at the behind of the test bed. The completeness of the generated operation program was proven by applying it to the conventional assembly station. The boundary conditions, such as the position tolerance range within 5~10 mm, were fulfilled.

5. SUMMARY

In this paper, the AR system which can be used at the manufacturing system was developed based on general software development method. The fundamental functions as the video interface, the tracking and rendering were implemented. And the manufacturing field condition-oriented auxiliary functions were added.

The function of generating clipping plane and controlling for precision and accurate work was used effectively during generating robot operation program. And the threshold value control function has similar effectiveness. These results attest functionality of the developed AR system which has high usability when the manufacturing system configuration is performed.

The function of the distance measurement between the virtual objects, and the dynamic movement function of the virtual objects are required after this research. Also, not only digital video camcorder but also CCD camera which is used generally in the manufacturing industry needs to be able to use with AR system.

ACKNOWLEDGEMENTS

This research was supported by the MKE( Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Assessment) (IITA-2008-(C1090-0801-0039))

REFERENCES

Design of Collaborative BOM Management System for Small and Medium Enterprises

Bo Hyun Kim*, So Young Jung, Jae Yong Baek, Seok Woo Lee, and Hun Zong Choi

Digital Collaboration Service Center
Korea Institute of Industrial Technology
Ansan, Kyunggi, 426-171, Korea

ABSTRACT

Most commercial PDM (product data management) systems, which have been released recently by overseas famous software vendors, are still expensive and heavy to apply to small and medium enterprises (SMEs) organizing a cooperative consortium. Proposed in this paper is the architecture and functions of a light and collaborative design management system called collaborative BOM (bill-of-material) management system (cBOM-MS). The cBOM-MS consists of BOM navigator which is a client/server based system and supports core functions to manage BOM in each company and collaborative BOM (cBOM) portal which is a web based program and provides the consortium with completing the collaborative BOM cooperatively. In the cBOM-MS, BOM Navigator and cBOM Portal are positioned separately to reinforce the information security in the consortium during product development process.

1. INTRODUCTION

Nowadays, the global business environment tends to change to the direction that not only the customers’ diverse demands are swiftly reflected to the products to preoccupy the market but also their latent demands are predicted to create a new market. This environmental change strongly requires the innovation and change in the value chain of the product development as well as the product itself. But, as small and medium enterprises (SMEs) have no sufficient internal innovative capacity, they have to cope with this situation through the collaboration within the value chain of product development. That is, the small and medium manufacturers must endeavor to construct more diverse collaboration spaces in order to escape the simple OEM system and develop their own collaborative products.

In the product development procedure, diverse product-related data are generated and diverse transactions such as the collaboration with the associates are executed. Therefore, the systematic management is necessary in this space. Product data management (PDM) is a concept to obtain the total efficiency of engineering through the integrated management of all business process and information in the product life cycle. In general, a PDM system includes diverse functions to manage the product-related data such as project-related data, technical documents, drawings, and images and the related businesses process. Thus, it tends to be complex and expensive.

The collaboration among enterprises may occur in every product development process such as product planning, design, and manufacturing. In general, most SMEs in the value chains participate in the collaboration naturally. In the view point of individual enterprise, it participates only in a part of the collaborative project. Instead of introducing the expensive PDM system, thus it is desirable to use lighter and cheaper system which manages the basic information of the product clearly and efficiently. World-famous S/W companies have released diverse PDM systems with high functions, but it is still difficult for SMEs to purchase them in the aspect of use and expense due to this.

For the collaboration among SMEs in specific areas of the product development process, some web-based collaborative systems began to be developed in around 2000 to support the information sharing and communication among the subjects of planning/design/manufacturing who were geographically separated [1,2]. After then, the conference function which enables the data sharing and chatting via a viewer was developed [3,4], and diverse application technologies for the collaborative design were also proposed [5].

* Corresponding author: Tel.: (0082) 31-436-8054; Fax: (0082) 31-436-8050; E-mail:bhkim@kitech.re.kr
This study proposes the structure and functions of a design information management system centering on the BOM (bill of material) which can support the collaboration among SMEs in the product development process. The proposed system has expanded its functions to express the information of diverse configuration specifications on the basis of BOM function and the traditional basic information. In addition, it supports the collaborative environment, has enhanced the information security function, and was designed lightly to be fit for SMEs. Of the existing studies on BOM system, there are the Generative BOM processing system which supports the efficient management of BOM for a lot of end-item specifications [6-9], the Procedure-Oriented BOM system based on the structure of programming language [10], and the Variant BOM system using the coordinate concept [11].

2. USER REQUIREMENT OF COLLABORATIVE BOM

Before describing the user requirements of collaborative BOM (hereinafter, cBOM), let’s survey SmarTeam® and TeamCenter®, the commercial PDM systems based on CATIA® and UG® respectively, which are widely used in the automobile industry. SmarTeam® is the PDM system of IBM. It accumulates the data generated during the design, manufacturing, and maintenance stage and integrates the processes through the automated workflow and the management of revised items to provide the wide-ranged security, revision, document link, structure management, and decentralization environment for all product information [12]. On the other hand, TeamCenter®, the PDM system of Siemens, provides the efficient configuration management and variant management functions integrating total product structure, collaborative CAD assembly modeling, top-down design, and structural digital mock-up functions. Moreover, it shares the design model in the workflow-based process and supports the collaboration in the perfect digital environment [13]. Both of the systems are covering the whole period of the product and manage the product data through the workflow-based process integration. But the wide-ranged management function for all the products and documents is still too heavy and expensive for SMEs.

One of the most important functions of PDM is the management function of product structure and configuration. The product structure shows the relation of up-down assembly among the parts and has the information such as upper part, lower part, and necessary quantity. The product configuration shows the variant information such as the assembly specification and options. Only when this information is efficiently managed, the PDM can be used appropriately in the industrial site [14].

The cBOM management system (cBOM-MS) proposed in this study is to be designed lightly for the easy use by SMEs and to provide the product structure and configuration management function, the essential function of PDM. The cBOM system supports the product structure and configuration management on the basis of BOM. BOM is the central information of the product, and is used not only in the design but also in the manufacturing and the maintenance. Through the interview with design managers of SMEs, the requirements of cBOM-MS were found out as follows:

(1) Link of BOM-drawing-technical document: If BOM presents the relation among components and the characters of components, the drawing shows the concrete details of product and its components. The drawings express not only the image information but also the dimension information, the material information, and the processing information. The information not expressed on the drawing such as standards, environmental regulations, test conditions, and requirements of each country is described in the separate technical documents. For the easy and swift use in the project site, the basic information of BOM should be linked with the drawing information of each product and its components on the BOM list, and the related technical documents should also be checked together.

(2) Separation of internal-external collaboration for technical security: As the drawings and technical documents related with the product development are the essential asset of the company, the security management is very important during the internal and external collaborations of the company. Even if the collaboration is made, the existing processes

![Figure 1: BOM-centered link of CAD-technical documents](image-url)
and the product technologies of each company are the private asset of the company and should be protected. Meanwhile, as the added values generated in the collaborative space are the common asset of collaborative system which does not belong to one particular company, they should also be managed and protected in common. To satisfy these two contradicting requests, the BOM system used inside the company should be separated from the external collaborative system for the inter-company collaboration. Of course, the security management is possible even in the technically integrated system. But the personnel in charge of product development in the company strongly require the physical separation of the system.

(3) Provision of collaborative design function: Collaborative product development is neither the simple outsourced production nor the one-sided transmission of specification but the cooperative development with the joint development process. Thus, the cBOM-MS should provide designers of SMEs with the exchange of BOM as the basic design information, the delivery of drawings for the relevant product and the parts, the sharing of BOM structure and configuration meanings, and the management of revision during the collaboration of product design.

(4) Expansion of collaboration range: The collaboration range in the enterprise activities is indescribably diverse. The requirements and characters of collaboration space are all different according to the market size and the characteristics of the product. Thus, to be used easily by small and medium manufacturing enterprises, the proposed system should be executed easily and intuitively in the input of basic information and the definition of drawing number system for the flexible management of design information regardless of specific market or product group.

3. STRUCTURE AND FUNCTION OF cBOM SYSTEM

As mentioned in the previous chapter, the users strongly require the separation of internal and external collaborative business because of the security of technology. That is, as the essential technical information is stored and managed in BOM, it is natural to think of the security as important. In fact, the collaboration between companies is generally the vertical-type cooperation. Here, a leading company shares the minimum information for the collaboration and keeps the rest information secure strictly. In this study, the cBOM-MS is composed separately of the BOM navigator which can be operated inside the company and the BOM portal for the collaboration with other companies as shown in Figure 2. The BOM navigator provides the BOM interface function for the connection with general commercial CAD systems, and the BOM portal becomes the place to exchange the BOM among companies.

Figure 2: Structure of cBOM management system(cBOM-MS)
3.1. BOM NAVIGATOR

BOM navigator is the client/server-based BOM management system for the internal collaboration of the company, providing the information communication between design department and related department and the management of product data. In this study, the BOM managed by the BOM navigator inside the individual company is defined as the company BOM. Especially, as the company BOM generated by the leading company in the collaborative consortium has influenced on the total collaborative projects and is used as the standard, it is defined as the standard BOM. That is, the standard BOM is the basic BOM in the collaborative consortium, and the leading company and the cooperative companies should make the design cooperation to complete the reference BOM. The cBOM is the part (module/parts) separated from the standard BOM for the actual collaboration with the cooperative companies.

For example, let's suppose the situation that an automobile module is developed in the collaborative environment as shown in Figure 3. Here, the leading company is the manufacturer of the module, and at this time the module is consisted of two sub-modules as shown in Figure 3-(a). For the sub-module 1, the sub-module 1 composed of two parts is designed by the leading company itself as shown in Figure 3-(b). Meanwhile, the detailed designs of part 2-1 and part 2-2 of the sub-module 2 are carried out by the cooperative company 1 and 2 respectively as shown in Figure 3-(c) and (d). The part 2-1 has the specifications of ‘E’ and ‘F’, and the part 2-2 has that of ‘G’ and ‘H’. Here, the standard BOM is the BOM of the leading company, and the sub-module 2 including the part 2-1 and 2-2 defined for the collaboration of cooperative company 1 and 2 becomes the final cBOM as shown in Figure 3-(e). The final standard BOM is completed by combining the BOM of the sub-module 1 with the final cBOM as shown in Figure 3-(f).

![Figure 3: Completion procedure of the standard BOM through the collaboration](image)

The functions of the BOM navigator are largely classified into the regions of the company BOM management and the cBOM management. Each region consists of the main functions such as the management of BOM structure and configuration, the BOM edition, the auto numbering, the management of design revision, the management of history, and the link management of BOM-CAD-Technical documents. As shown in Figure 4, the company BOM has the
Design of Collaborative BOM Management System for Small and Medium Enterprises

hierarchy of components of a product or a module, and each component is composed of its registration information, history, related drawings, and related technical documents. In other words, the BOM tree presents the whole information of a product/module/part including the assembly structure of its components, its registration, history, related drawings and documents. From this reason, we can easily extract the hierarchy of drawings and technical documents independently from the BOM tree.

![BOM tree of a product/module/part](image)

Figure 4: BOM tree of a product/module/part

According to the various demands of customers, a product/module can be presented in several models, and even in one model which can be various variant products/modules according to the specification. Because of this, the information of products to manage grows in geometric progression, and the data overlapping problem can seriously occur in the aspect of data management. To solve these problems, this study proposes that the individual parts are maintained with the structure information and the variant information, and this information is expressed on BOM. In the traditional BOM systems, the basic view expressed in the form of tree is presented to manage the up-down relation between parts as well as the structure and the quantity. Meanwhile, the configurable view as shown in Figure 5 is provided to express the variant information in the proposed BOM management system. Furthermore, when the products, or modules, or parts are registered, the attributes of variant information are added/deleted, and the related information is presented in the relevant BOM according to the attributes and level of variant option.

![An example of the configurable view of a product/module](image)

Figure 5: An example of the configurable view of a product/module

<table>
<thead>
<tr>
<th>Level</th>
<th>SEQ</th>
<th>Part name</th>
<th>QTY</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>SEQ</td>
<td>Part name</td>
<td>QTY</td>
<td>PVC</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>A/T SHIFT LEVER</td>
<td>1</td>
<td>PVC</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>KNOB A RHD</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>KNOB PVC</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>KNOB LTR</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>BUTTON</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>KNOB A LHD</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>KNOB PVC</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>KNOB LTR</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>BUTTON</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>Housing 5S</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>Panel 5S</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>BRKT</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>Panel 4S</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>BRKT</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
The cBOM-MS for the external collaboration is the functions to generate and manage the cBOM by revising and editing the part of BOM of the leading company. That is, these functions are composed of the management of the registration information, history, related drawings, related documents, and input of the cooperative companies. For example, from the standard BOM of the target product/module some module/part to be designed collaboratively is separated to form the initial cBOM. Next, the information of relevant cooperative companies is inputted into the separated modules/parts and registered in the cBOM portal. In the cBOM portal, the registered cBOM is embodied through the collaboration with the cooperative companies. The embodied cBOM is verified and finally combined to the standard BOM.

3.2. COLLABORATIVE BOM PORTAL

In reality, the software module to execute the collaboration between companies is the cBOM portal. The cBOM portal exchanges the registered cBOMs with each other and manages the progression schedule of BOM, all the histories related with the cBOM, and the authority and the security of each company. The cBOM which is first registered on the cBOM portal, the initial cBOM, has not the detailed design of its components (modules/parts) but the specification of them. While the collaboration is progressed, the detailed design for each module/part is made, and the final cBOM is completed as the BOM tree of relevant module/part including the detailed design and related technical information. On the cBOM portal, each company can only register and read the cBOM. That is, unless the collaborative consortium allows the additional authority to it, the relevant company can only read the registered data and cannot delete or revise the data. Moreover, as all the records including the upload and download can be checked through the history management, it is clearly known when and who have used the system.

For example, in case of Figure 3, the leading company selects the sub-module 2 to execute the collaborative design using the BOM navigator and generates the initial cBOM from its BOM. In the initial cBOM, the sub-module 2 consists of part 2-1 and part 2-2. It has the specification information in which the part 2-1 and part 2-2 have the specification of E/F and G/H respectively and the information of cooperative company 1 and 2. Next, the leading company registers the initial cBOM on the cBOM portal to allow the cooperative company 1 and 2 to execute the detailed design. At this time, the initial cBOM is opened only to the cooperative company 1 and 2 to execute the collaborative design. If each cooperative company registers its BOM including the detailed design specification, drawings, and technical documents on the cBOM portal within the due-date, the final cBOM is completed. The leading company can read and use all the information of final cBOM including the opened information of the cooperative companies. However, the cooperative company can read not the information of other cooperative companies but the registered information of its own and the information opened by the leading company. After download of the final cBOMs to the BOM navigator, the leading company makes the final examination on them and completes the total standard BOM. This sequential BOM exchange process centering on the cBOM portal is explained in Figure 6 in detail.

![Figure 6: BOM exchange process](image-url)
### 3.3. BOM INTERFACE

Most companies create drawings using commercial CAD systems. In general, as the commercial 2D CAD systems usually express the product image in a planar way, the BOM should be separately generated. But recent 3D CAD systems can express the product information and the structure information of configuration parts in a tree form. Thus, for the better use of the proposed BOM navigator, it is necessary to provide the function to directly extract the BOM-related information created by the commercial 3D CAD system and to draw out the BOM automatically. In other words, the automatic BOM generation function is the interface function between BOM navigator and 3D CAD systems. This study developed this interface function and carried out its functional test in the CATIA® and UG® systems which are actively used in the automobile industry.

### 4. CONCLUSION

This study proposed the structure and the function of a BOM-centered design information management system which can be used for a SME consortium to carry out the collaborative development of new products. The proposed system has the structure and the function to manage the design data in connection with the part/drawing/technical document centering on the BOM which becomes the standard information of the product structure and configuration management. Especially, as the design information is an important asset of the company, the vulnerable security function in the collaborative environment was overcome by linking and separately operating the BOM navigator for the internal collaboration of the company and the cBOM portal for external collaboration among companies. In addition, as an interface to maximize the usability of proposed system, the automatic BOM generation function from the commercial 3D CAD system was proposed.

In the BOM navigator, the configuration management view was provided along with the basic view for the organic structure and specification management of products, modules, and parts and the variant information management. The company BOM of the leading company who leads the collaborative consortium is defined as the standard BOM, and the cBOM is drawn up by editing a part of standard BOM in the collaborative region and is open to the cooperative companies through the cBOM portal for the collaborative design among the companies. In the cBOM portal, the security by companies is managed so strictly that each cooperative company can read and revise only the information opened by the leading company and registered by the cooperative company itself. The proposed system is in the stage of prototype implementation as shown in Figure 7 and will be stabilized and complemented through the application to the design collaboration sites in the near future.

![Figure 7: An example image of prototype cBOM-MS](image)

### ACKNOWLEDGEMENTS

This research was carried out as a part of i-manufacturing (Korean type innovation of manufacturing) project by the Ministry of Knowledge Economy, to whom we give thanks.
REFERENCES


Active Disassembly application to enhance Design-for-Remanufacturing

Joseph D Chiodo and Winifred L Ijomah

1Active Disassembly Research Ltd.
London, UK

2DMEM
University of Strathclyde, Glasgow, UK

ABSTRACT

Developing sustainable approaches to remanufacture is a critical global concern because of manufacturing’s significant adverse environmental impacts. Remanufacturing is the process of returning a used product to at least original performance specification from the customers’ perspectives with at least “as new” product warranty. The practice is a more sustainable manufacturing mode because it is profitable but less harmful to the environment than conventional manufacturing. This paper discusses how ‘Active Disassembly’ (AD) could extend the application of remanufacturing in sustainable manufacture and proposes areas for further research. AD enables product non-destructive self-disassembly at end of life and was invented to facilitate a step-change improvement in recycling. However, the rapid disassembly with reduced component damage that it enables can increase remanufacturing economic viability. This is the first investigation into this specific link.

1. INTRODUCTION

Major global concerns include competition and environmental pressures, in particular waste from manufacturing. Significant factors of improvement must therefore be made in industrial practice and technological solutions [1] in order to attain a more sustainable mode of production and consumption. Product design must reflect this [2, 3]. Current product LCA demonstrates that the disposal phase contributes substantially to the environmental impact of Waste Electrical and Electronic Equipment (WEEE) [4, 5, 6]; particularly in products containing toxic materials [7], scarce or valuable materials, or materials with a high energy content. Within WEEE there is the combination of all three of these situations including for example, batteries, quality plastics, precious metals, and toxic solder. 1998 figures of E.U. WEEE was in the region of 6.5 to 7.5 million tonnes per year [8, 9], representing <1% of the total EU solid waste stream [10] and responsible for at least (base case scenario) 50% of metals in leachate, however, the quantity of electronic products and their waste is continuing to rapidly increase [11].

Remanufacturing, a process of returning a used product to at least original equipment manufacturer performance specification from the customers’ perspective with equal new product warranty [12, 13], is a central and profitable strategy in waste management, material recovery and environmentally conscious manufacturing. Remanufacturing’s significance is that it limits waste generation as well energy and resource consumption by manufacture. In contrast, conventional manufacturing generates in excess of 60% of annual non-hazardous waste [14]. Furthermore, increasingly severe legislation, such as landfill and end-of-life directives, demand reduction in the environmental impacts of products and manufacturing processes [15]. Since the Basel agreement [16] prohibits exporting of waste outside the EU, European producers must manage their waste internally. This, combined with rising fiscal penalties such as the landfill tax [17], and increasing global competition, makes remanufacturing expertise paramount. A key problem is designer lack of expertise in designing remanufacturable products [12, 18]. Currently remanufacturing is profitable typically for large complex mechanical and electromechanical products with highly stable product and process technology [12, 18]. Research indicates few examples where products with volatile product and process technologies are routinely remanufactured, chiefly as their remanufacture is cost prohibitive. ‘Active Disassembly’ (AD) is an alternative to conventional dismantling techniques that enables the non-destructive, self-disassembly of a wide variety of consumer electronics on the same generic dismantling line, thus reducing disassembly cost [19]. This paper illustrates the potential of applying this technology to WEEE remanufacturing. The importance here is that WEEE is the EU’s fastest growing waste stream [20], but small sized WEEE are typically not profitable to

* Corresponding author: Tel.: +44 (0)141 584-4346; E-mail: w.l.ijomah@strath.ac.uk
remanufacture as their volatile technological pace makes their disassembly by conventional means overly expensive. Remanufacturing aided by AD seems a natural progression of technology cross-fertilization.

2. Methodology

Remanufacturing related findings were obtained via literature search, telephone interview and observational case studies supported by interviews with key company personnel, informant examination of case study reports and experimental analysis. Results validation was by stakeholder review and peer review via publications. The stakeholder review involved testing for replication logic [21] by discussing the results with new remanufacturing practitioners at trade fairs and conferences as well as telephone discussion with new remanufacturers. For AD, as first conceived and implemented in 1990 by Chiodo, the technique has been applied to a variety of electronic products. Various applications and variations by other researchers and companies have contributed to this increasing body of knowledge [22, 23, 24, 25, 26, 27, 28]. Therefore, results of these findings were applied to design guidelines for remanufacturing such as those by Ijomah, [29]. Examples of candidate AD products include mobile phones, a variety of other small hand held electronics, home entertainment goods and automotive electronics (ECUs).

3. The remanufacturing domain

The key differentiating factors between repair, reconditioning, and remanufacturing include: (1) remanufactured products have warranties equal to new alternatives whilst repaired or reconditioned ones have inferior guarantees. Typically, with reconditioning the warranty applies to all major wearing parts whilst for repair it applies only to the repaired component. (2) Remanufacturing generally involves greater work content than the other two processes leading to products that generally have superior quality and performance (3) Remanufactured products lose their identity whilst repaired and reconditioned products retain theirs. Furthermore, remanufacturing requires assessment of all product components with those that cannot be brought back at least to original specification replaced with new (4) Remanufacture may involve upgrade of used products beyond the original specification which does not occur with repair and reconditioning. Reuse is the process of using functional components from retired assemblies [30]. Recycling is the series of activities for collecting, sorting and processing discarded materials for use within new products and is currently far more common than remanufacture [31]. Remanufacturing is preferable to recycling because it adds value to waste products by returning them to working order where as recycling simply reduces the used product to its raw material value [12].

![A generic remanufacturing process flowchart (Ijomah, [12])](image-url)
Remanufacturing begin with the arrival of a used product (a core) at the remanufacturer’s facility where they undergo a series of activities: disassembly, cleaning, component remanufacturing, replacement of unremanufacturable parts, reassembly and testing to produce the remanufactured product. All remanufacturing operations have this basic structure but the order in which these activities shown in Figure 1 and described in [32] are undertaken may differ between different product types [33]. Evidence of the environmental and economic benefits of remanufacture is well recorded in literature [12, 18, 34]. As consumers purchase remanufactured goods to obtain high quality products at low prices, remanufactured products must have high quality, reliability and low price, to compete against alternatives such as reconditioned and new products [12]. Much of the work to enhance products’ potential for remanufacture has been concerned with developing design-for-remanufacturing guidelines (DFRem) to assist designers because they lack remanufacturing knowledge. Examples of research in this area include [35, 36, 37, 38, 39, 40]. In addition to the above work that look specifically at remanufacture there are also other research that although not totally geared towards remanufacture none the less produce guidelines that can enhance remanufacturability, for example, [41, 42, 43, 44]. Although these research findings are commendable they do not address the issue of extending remanufacturing to product that experience high rates of technological change. The following sections will describe the AD concept and its potential to extend remanufacturing to fast moving consumer products.

4. ACTIVE DISASSEMBLY (AD) DOMAIN

AD, is a process by which an EoL product can non-destructively, self dismantle with the aide of embedded smart devices that are triggered by external energy input and are designed as part of the manufacturing and controlled hierarchical disassembly of the host product. In this study, three types of materials are employed within AD; the first two being either metallic or plastic. This process includes NiTi or Cu-based ‘Shape Memory Alloys’ (SMA) for AD-SMA, used for actuator devices. Polyurethane based ‘Shape Memory Polymers’ (SMP) are used for AD-SMP releasable fastener devices. Actuation of the SMA and SMP would occur just outside of world ambient temperatures for safety and practical reasons and would only change shape under the predetermined temperature. This ‘predetermined’ temperature is based on material composition and is therefore consistent and stable. This trigger temperature is referred to as ‘Tx’. The third ‘material’ is a conformal coating spray or ‘interstitial layer’ (IL); hence AD-IL. AD is particularly useful for complex assemblies typically with disassembly problems such as the requirement for very high product environmental performance (for example, extreme temperatures or vibration). Using SMA in the design of actuators, it is possible to effectively force the fastening elements of the products apart. This happens at the composition specific SMA Tx range; Austenite starting to finishing temperature (As-f). Below As the SMA actuator device lies dormant until exposed to it’s full Tx. SMA therefore possesses the ability to be stable in two different temperature states. When used in the design of a ‘smart material device’ (SMD), SMA is a force provider when triggered at this very narrow temperature bandwidth As-f, typically 14degC. SMA can be one-way [45], two-way [46] and less often, multi-way when used in the design of an actuator. SMP acts differently as it immediately looses shape integrity over a small temperature band with no accompanying force. It is therefore useful in the design of releasable fasteners capable of significant metamorphic shape change. SMP can be ‘called’ when triggered within a specific temperature exposure and range [47]. This is accompanied by a sharp drop in modulus. This engineering feature allows for an effective disassembly when the gripping portion of the SMP fastener looses its ability to hold its host together at its composition specific Glass Transition Temperature (Tg). When employed by AD, this would occur within its full temperature range completed at Tx. Just like SMA, the SMP SMD, lays dormant within its host assembly during the product’s use phase waiting to be triggered at EoL. Besides shape memory materials, there are numerous other ‘smart’ materials that can be employed to facilitate AD. These include engineering polymers, smart films, bio-degradable layers, adhesives, substances, liquids [48]. The technology offers a variety of dismantling applications. Numerous self-disassembling techniques have been investigated with varying levels of success, [22, 23, 24, 25, 26, 27]. It has been considered that these approaches can be categorized as automatic processes but labelled as ‘product embedded disassembly’ [28]. Different forms of AD also exist [49]. The following sections describe the application of AD to optimise WEEE recycling.

5. APPLICATION OF ACTIVE DISASSEMBLY TO WEEE RECYCLING

In conventional disassembly EoL products are dismantled manually, typically using pneumatic or hand tools. Manual disassembly has been used in recycling plants around the world to varying levels of success since the early 1990s. With high turn-over life cycle products like those in the technology sector, manual disassembly hinders profitable disassembly for telecommunication or other hand held electronic products. Significant economic

* Corresponding author: Tel.: +44 (0)141 584-4346; E-mail: w.l.ijomah@strath.ac.uk
efficiencies in disassembly is thus required to assure competitiveness in disassembly oriented EoL scenarios [28]. Invented to facilitate a step-change improvement in recycling, AD aims include hierarchical, controlled, non-destructive and specific product component release from macro to micro dismantling in order to optimise component re-use. The technology has been applied with great success in numerous items from automotive electronic control units (ECUs) through to laptops to home entertainment products, their sub-assemblies and most un-bonded components. Constituent assemblies here include housing, antenna, PCB, transformer, ECU and mechanical components. Figures 2, 3, 4 and 5 show the successful testing for non-destructive disassembly of a variety of WEEE products prepared with AD technology.

![Figure 2: Plastics dominated WEEE: AD-SMP embedded home entertainment (Chiodo et al. 2002 I Mech E)](image1)

![Figure 3: Precious metal dominated hand held WEEE: AD-SMA embedded (Chiodo et al. 2002 JM&D)](image2)

![Figure 4: Precious metal dominated WEEE: AD-SMP mobile (Chiodo 1998, ADR Ltd.)](image3)

![Figure 5: Metals/precious dominated automotive WEEE embedded with AD-IL (ADR. Ltd. 2002 and Arnaiz et al. 2002)](image4)

The tests here include hierarchical non-destructive disassembly which is particularly useful when the re-use required is of selected parts as opposed to all product components. These products, ranging from hand-held electronics to larger automotive electronics were modified with embedded AD-SMP, AD-SMA devices or an AD-IL. The figures illustrate various before and after samples. Figure 2, shows the pre AD of a Sony Playstation using an AD-SMP screw. SMA was used for actuator applications for AD in Figure 3. Precious metal content goods, as seen in Figures 4 and 5, illustrate potential use of AD with mobile phones and automotive goods also with high amounts of quality metal, aluminium in this case). Three different AD technology variants were tested and all examples were successful with non-destructive and low mean time dismantling. In addition, Chiodo has invented and experimented with numerous alternative approaches of AD technology to expand the potential repertoire of candidate products, for example, automotive ECUs as illustrated in Figure 5. This was employed because ECUs are almost impossible to economically disassemble cleanly. AD disassembly times ranged from one to 100s of seconds.
on products not redesigned for AD inclusion. Exact times for this study are irrelevant since any cost advantage in dismantling would only likely be realized in batch processing [19] as per any automated economically viable system. These new developments by Chiodo have illustrated the potential for more efficient and step change approaches to disassembly friendly product dematerialization design and manufacture strategies [48]. For the highest rate of automatic dismantling control: hierarchical, non-destructive, lowest mean time; would be achieved only if the products were designed with AD from the outset. Factors to be considered include a number of guidelines described in [50] and shown truncated in Table 1 below.

<table>
<thead>
<tr>
<th>1. Disassembly methods:</th>
<th>Tx, trigger methods, mechanical aide (tumbling etc.), fastener, deflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Active material choice:</td>
<td>force providers (actuators), modulus drop, dissolving, magnetic, mechanical</td>
</tr>
<tr>
<td>3. Actuator design:</td>
<td>force control, contact, pressure, location, exposure, deflection, integrated, discrete</td>
</tr>
<tr>
<td>4. Product design:</td>
<td>deflection, release allowance and passage, trigger exposure, re-use, DfE</td>
</tr>
<tr>
<td>5. Orchestration:</td>
<td>hierarchy, Tx/time, assembly arrangement,</td>
</tr>
<tr>
<td>6. Use phase:</td>
<td>Tx and trigger range outside of ambient exposure, creep, reliability</td>
</tr>
<tr>
<td>7. Output logistics:</td>
<td>volume, time, cost, quality, batch throughput (product types)</td>
</tr>
<tr>
<td>8. Recycling:</td>
<td>markets, environmental assessment</td>
</tr>
</tbody>
</table>

### 6. POTENTIAL FOR USE OF ACTIVE DISASSEMBLY TO ENHANCE WEEE REMANUFACTURING

Disassembly is typically an initial and critical activity in remanufacturing thus long disassembly time raises remanufacturing cost and lead times. Research, for example [12, 51]; have shown that the ability to disassemble is key for remanufacture because of its significant impact on the remanufacturing success factors of price, delivery and quality. In the case of price, customers are unlikely to purchase a remanufactured product unless it is less than ½ the price of the new alternative. However, high remanufacturing costs often result from resource consuming disassembly due to the time (and hence) labour costs and the need to replace components damaged during disassembly. Secondly, customers are unlikely to buy a poorly functioning product no matter how low its price. However, sound remanufacturing typically requires effective disassembly to ensure proper cleaning, assessment and rectification of internal components [12, 32]. Thirdly, unless the price of new products is prohibitive customers will purchase new alternatives rather than wait for lengthy time periods for a remanufactured version to be available. The above case studies clearly illustrate AD’s benefits in ensuring non-destructive disassembly coupled with reduced disassembly time and cost and hence it’s potential to enable profitable remanufacture of products that are currently not economically viable for remanufacture. Hand held WEEE including PDAs, calculators and mobile phones have been tested since 1996, originally for precious metal fraction reclamation [53], but there is also the potential for remanufacturing markets if enough through-put is achieved. Plastics dominated products such as home entertainment and brown goods also show promise, as illustrated in Figure 2. In assemblies where high performance sealants contaminate the potential for economic remanufacture, an automatic process is required to, through triggers including temperature (Tx), vibration and shear forces (tumbling), cleanly remove the sealant from the assembly as described in [48] and [54]. If products are designed for AD, time reduction and minimal (if any) evidence of dismantling would be achieved to increase the economic viability of their remanufacture.

### 6.1 BASIC FINANCIAL ANALYSIS

Figure 6, shows three categories of products of interest to implement AD within a pre-remanufacturing scenario; all but CRT products. For cost reduction to be viable, AD must be performed in large batch disassembly prior to remanufacture. Previous studies indicated that likely batch sized would be from 500 products to 1000 per 10 seconds [19]. Metal dominated products were not previously seen as viable, but the ECU presents a more positive scenario due to the high grade Al % content and a large PCB. Since the increase in automotive electronics is expected to continue AD presents the opportunity for 25% minimum cost reduction required for their market viable remanufactures. The following sections assess the potential for viable remanufacture of plastics dominated, common metals and precious metal dominated ELV WEEE products when AD-IL is applied. This process has an additional manufacture cost: the interstitial layer application: [48] ( http://www.activedisassembly.com/animations/il.html). However, it requires minimal design changes and provides clean separation from adhesives, silicon and other setting bonding agents that are typical in automotive ECU applications requiring high performance against the elements. When such products are redesigned with AD-IL inclusion, a batch of 100 to 1000 products could dismantle on average, every 10 seconds. Here, there is the potential to process five batches per minute in a large scale optimized
Remanufacturing is the process of returning a used product to at least its original performance from the customers’ perspective and giving it a warranty at least equal to that of a new product. It is an important re-use strategy in waste management, material recovery and eco-conscious manufacturing because it is environmentally and economically beneficial. AD is a process by which an EoL product can rapidly and non-destructively self dismantle with the aid of embedded smart devices to optimize component reuse. The major remanufacturing barriers include poor understanding of how to design products for remanufacture. Disassembly is typically an initial and critical activity in remanufacture. However, the research has shown that many products are not remanufactured because they cannot be non-destructively disassembled or their disassembly is expensive. This severely limits their potential for remanufacturing because remanufactured products must be at least 25% cheaper than new alternatives to win customers. Remanufacturers rejected jobs that were beyond economic repair (BER) unless there were strategic reasons, where BER was typically assessed as costing above 70% and 65% of new cost. Given the proven success in AD application to optimize component re-use in WEEE recycling, it seems natural to marry AD to design-for-remanufacturing. To this end, the rapid disassembly with reduced component damage that it enables results in increase in remanufacturing feasibility to extend the range of products that can be cost effectively remanufactured. Further research would include firstly, the application of AD to the design of selections of WEEE products to ascertain and quantify the impact on their remanufacturability and, secondly, the assessment of the viability of a market for the remanufactured, re-designed WEEE products.

Acknowledgements

The authors wish to thank Active Disassembly Research Ltd., EnviroTech Ltd., Motorola, Nokia, Sony, Gaiker, Indumetal, IKP, Active Recycling Ltd. and Brunel University, ATP Industries Ltd, J and E Hall, Rolls-Royce Plc, Caterpillar Remanufactured Products, DML Devonport, Lec refrigeration Plc, Ivor Searle Ltd, Price Brothers Auto.
REFERENCES


The Fault-Ride-Through Capacity of Doubly-Fed Induction Generators (DFIGs) for Wind Applications

Wenping Cao*, Jessada Stevenson

School of Science and Technology
University of Teesside
Middlesbrough, England, TS1 3BA

Abstract

The behaviours of the grid connected Doubly-Fed Induction Generator (DFIG) responding to grid disturbances are studied in this paper. Grid disturbances include voltage sags resulting from symmetrical and unsymmetrical short-circuit faults. The analysis is carried out using real-time simulations on MATLAB’s SimPowerSystems models. Two types of faults including the single-phase and three-phase faults are simulated on a 3.96 MW wind farm connected to a solid 150 KV network feeder. A crowbar power electronic protection scheme is implemented in the circuit and used to aid the fault ride-through of the DFIG.

1. Introduction

Wind turbine utilisation is seen in rise in the electricity generation since the wind energy represents a promising source of renewable energy. As a result, numerous studies have been conducted to maximise the wind energy capture or to optimise the control strategies. However, the integration of wind energy into the power grid imposes certain impacts on the power grid, particularly the stability issue of the power system. On the other hand, the immunity of a wind farm to faults on remote and local buses from the power grid can also be of interest. The latter is often referred to as Fault-Ride-Through (FRT) capacity and is the focus of this paper. Figure 1 shows fault ride-through profiles for some European countries.

Despite early stagnation, wind power has grown very fast and has become commercially competitive with other means of electricity generation. At present wind turbines are being used extensively for large-scale deployment and play an increasingly significant role in the power network. Figure 2 illustrates the installed wind capacity for the year 2007. The UK, as the richest country in Europe in terms of wind energy potential, has installed only a small fraction of the world’s total wind turbines. This is partially due to the limited locations available for onshore wind power generation. Therefore, a significant growth area is expected to be offshore wind farms in the UK.
In the past, with low penetration of wind power into the grid, the power system was not affected by the presence of the wind farm. That is, the loss of power generation from a small wind farm did not pose a threat to the security of the overall power system. Nevertheless, since the size of wind farms is remarkably increasing, and so is the penetration of wind energy in the power network. Although the power system technology has been progressing dramatically during the past two decades with the help of improved power electronic devices, the reliability and security of the power systems still remain a significant issue. Thus, a better understanding of the impacts of large wind farms on the grid stability is required before any wind farms are interconnected to the grid. This involves the use of advanced wind farm dynamic models and analytical methods in solving potential problems of grid stability.

In this paper, methods of predicting and improving the low voltage ride-through capabilities of the wind farm are developed by modelling the dynamic performance of a power system. Power system faults, including balanced three-phase and single-phase-to-ground are investigated, based on dynamic and steady state analyses. In the dynamic simulation, the dynamics of a 3.96 MW wind farm connected to a 150 KV feeder whose data are extracted from the literature [3] is studied with relay protection and circuit-breaker settings included. Simulations are carried out using MATLAB’s simpowersystems with both phasor and detailed models employed.

Since the Doubly-Fed Induction Generator (DFIG) is currently the dominant topology for large wind power generation, its behaviours responding to grid faults are investigated in steady and transient states. A parallel grid side rectifier and series grid side converter are used to remedy a possible ride through fault whilst keeping the DFIG connected during grid faults.

2. OPERATION OF DFigs

For wind power generation, there are various types of electrical machines can be used in parallel with the wind turbine. An overview of the different wind turbine generator systems and their comparisons is given in [4]. The variable speed concept with a partially scaled power converter is known as the DFIG configuration. This technology allows extracting maximum energy from the wind for low wind speeds by optimising the turbine speed, while minimising mechanical stresses on the turbine during gusts of wind. As shown in Figure 3, this induction machine consists of a conventional stator which is directly connected to the grid, and a wound rotor which is connected to the grid via power electronic converters through slip rings. Clearly, the presence of slip rings leads to some disadvantages as they would require regular maintenance [5]. Also in Figure 3, the rotor power converters includes a rectifier (AC/DC) and an inverter (DC/AC), with a capacitor linked in between acting as the DC source. Forced commutated high switching devices such as insulated gate bipolar transistors (IGBTs) are usually employed in the power converter set-up so as to ensure accurate AC voltage from the DC bus link. The primary purpose of the two power converters is to control the rotor frequency and hence the rotor speed. The typical variable speed range for DFIG systems is about ± 30% of the synchronous speed. The rating of the power electronic converters is approximately 25-30% of the generator rated power, which makes this configuration attractive from an economic point of view. Furthermore, the DFIG configuration can also adopt the pitch angle control of the wind turbine blades so that the incoming wind power can be stabilised when the wind speed is too high.
3. GRID FAULTS

In order to fully appreciate the DFIG’s response to grid faults, an understanding of the different grid disturbances and their occurrences in any power system is of importance. In fact, power faults from the grid are many. However, open-circuit faults are not considered in this paper due to their relatively low levels of fault current contributions. Short circuit faults can also be divided into symmetrical and asymmetrical faults depending on if they are balanced among the three phases viewed by a fault. In three-phase power systems the short circuit faults are ranked in order of occurrence as follows: single line-ground, line-line, double line-ground and three phase faults. As a result, in this paper only single line to ground and three phase faults are modelled since they represent the most frequent and the most severe types in practice [6], respectively.

3.1. SYMMETRICAL FAULTS

The term symmetrical fault refers to power system faults where all three phases are effected equally. Symmetrical faults include phase-phase-phase and phase-phase-phase-ground faults.

Symmetrical faults can occur when insulation fails due to system overvoltages caused by lightning or switching surges, by insulation contamination (salt, spray or pollution), or by other mechanical causes as a result of adverse weather including falling trees on transmission lines. The resulting short circuit current is determined by internal voltages of synchronous machines and system impedances between the machine voltages and the fault. Fault currents can be many times larger in magnitude than normal operating currents. If allowed to persist they can cause adverse damage to equipment connected to the power system. As a consequence, it is usually necessary to remove faulted sections of a power network from service as quick as possible.

When a balanced three phase fault occurs in a balanced system there is only positive sequence fault current as the positive, negative and zero sequence networks are completely uncoupled.

3.2. ASYMMETRICAL FAULTS

Asymmetrical faults represent a short circuit in one or two phases of a three-phase system. Asymmetrical faults occur more often than symmetrical faults.

Possible causes of asymmetrical faults are:

(i) Line-to-line short circuit between any two of the three phases when lines come into physical contact as a result of a broken insulator.
(ii) Line-to-ground short circuit between any phase and ground, often caused by physical contact as a result of weathering (lightning, gale, forces, etc).

(iii) Double line-to-ground short circuit between two phases when they come into contact with the ground as well as each other, commonly due to adverse weather.

4. Simulation Models

The simulations for this experiment are carried out on the Matlab's SimPowerSystems. The experiment is based on a 3.96 MW wind farm connected to a solid 150 KV network feeder. Its phasor model is produced using SimPowerSystems generic DFIG blocks and also includes the power IGBT electronic converters Rsc and Gsc, as is illustrated in Figure 3.

A schematic of the wind farm for simulation is shown in Figure 4 and the parameter values for this wind farm are given in Table 1. The reactive power is initially set at 0 pu and when the voltage regulation mode is adopted the reference is 1 pu. The threshold for voltage regulation mode can be modified and action can be taken when the voltage at the generator terminals fall below this threshold, i.e. a turbine trip can be adopted. In order to comply with the U.K grid codes [7], the turbine must ride through the grid fault. As a consequence, the operation of the DFIG and its steady state performance are studied first, followed by analysing its transient response to grid faults. In this case, a parallel grid side rectifier and series grid side converter are also included in the models.
### Table 1: Parameter values for the 3.96 MW wind farm

<table>
<thead>
<tr>
<th>COMPONENT</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network feeder</td>
<td>$U_{dc} = 150$ kV, $S_{dc} = 3000$ MVA, $R_e/Z_g = 0.1$</td>
</tr>
<tr>
<td>System transformer</td>
<td>$T_i = 50$ MVA</td>
</tr>
<tr>
<td>Wind farm generators (G1-G6)</td>
<td>6 x 660KW with variable pitch angle</td>
</tr>
<tr>
<td>Transmission line</td>
<td>$P_{dc} = 660$KW operating power factor 0.9.</td>
</tr>
<tr>
<td></td>
<td>$U_{dc} = 690$ V, $I_{dc} = 560$ A</td>
</tr>
<tr>
<td></td>
<td>$R_i = 0.215$ $\Omega$/km, $X_i = 0.334$ $\Omega$/km, $L = 10$km</td>
</tr>
</tbody>
</table>

5. RESULTS AND DISCUSSIONS

The detailed model for simulation is shown in Figure 5. The phasor model was first run to simulate a three phase to ground fault on the 20 kV busbar of the plant and then a single-phase to ground fault. The faults were simulated for up to 3 seconds. Any power system fault lasting more than 3 seconds without being isolated by system protection is a very severe and unrealistic scenario. Although it is better to implement a long duration fault, a maximum of 3 second fault cycle is indeed sufficient for this purpose.

![Figure 5: Simpowersystems detailed model of the system (with crowbar protection circuit)](image)

5.1. Symmetrical Faults

Two models were produced with a phasor model which employs a generic DFIG block connected to a distribution network and a detailed model which represents the DFIG and detailed power IGBT electronic converters (Rsc and Gsc). In the phasor simulation method, the sinusoidal voltages and currents were replaced by phasor quantities at the system nominal frequency (50 Hz). Obviously the phasor model only takes into account the positive sequence currents and voltages while the detailed model however uses sinusoidal voltages and currents. The reactive power was initially set at 0 pu (also referred to as reactive power neutral) when the phasor DFIG block is operating in var regulation mode. When the voltage regulation mode is adopted the reference becomes 1 pu. The threshold for voltage regulation mode can be modified and action can be taken when the voltage at the generator terminals fall below this threshold, i.e. a turbine trip can be adopted.
Firstly, a 3 phase fault was simulated on the 20 kV bus bar. The phase-phase-phase fault was simulated for 1 second during steady state. No fault resistance is involved since the fault was not to connected to earth. The wind-farm can be operated in voltage regulation mode with the reference value at 1 pu or in reactive power regulation mode with the reference value at 0 pu. The simulation was ran with the pitch angle controller set to 2°/s and then to 10°/s for comparison of the performance in damping rotor over speed disturbances. These values were chosen as they are the fastest acting pitch angle used in practice [8].

Figure 6 shows the response of the DFIG wind farm to the fault on the 20 kV busbar with reactive power regulation mode.

![Stator voltage (pu)](image1)

![Stator current (pu)](image2)

![Generated active power at the stator terminal (MW)](image3)

![Generated reactive power (MVar)](image4)

**Figure 6: Simulation results for a 3 phase balanced fault on the system**

From the simulation results of Figure 6, it is evident that the positive sequence voltage falls below any of the fault ride through profiles so that the system is unable to ride through this fault simulated. As a result, a protection scheme is needed to aid fault ride-through for the system.

The stator voltage dips to almost 0 pu as the Rsc struggles to maintain the voltage to its desired set point of 1 pu. The lagging pitch angle controller over damps the rotor speed after fault clearance which causes a reduction in generated power after fault clearance of 30s. The rotors over currents are significant rising the DC link voltage to 4 times its nominal value. However, the faster pitch angle controller is able to damp the over speed quickly and the effects of over damping are thus not significant. The effects of the over damping can be seen on both simulations results for 2°/s and 10°/s pitch angle controllers.

The DC bus link capacitor rises to a peak value of 5.25 kV, which is twice the rated value. This suggests high levels of surplus rotor current from the symmetrical fault simulations. Even at the fault instant the DC link voltage rises to approximately 2.5 kV, which is still above the capacitors capability.

### 5.2. Asymmetrical Faults

The phasor model was run to simulate a single phase to ground fault on the 20 kV busbar of the plant with the fault simulated for 3 seconds. Test results in var regulation mode are shown in Figure 7.

From these results, it can be seen that the variations in stator voltage and current, in active and reactive powers are all less significant compared to the three-phase fault, in particular the stator voltage and active power. Due to the nature of induction machine which draws reactive power from the grid, reactive power compensation is always required for DFIG applications so as to enhance the machine’s ride-through capacity. When the remedial method is applied, these variations are all reduced but the generators can ride through the faults defined in the grid codes.
In addition to the stator voltage, the rotor current is also a concern when evaluating the DFIG’s fault ride-through capability. The transient rotor current is largely dependent upon the power converter characteristics whose rating is only a fraction of the generator rating. When a disturbance occurs, there is a sudden step changes in AC stator voltage throughout the disturbance owing to the magnetic coupling between the rotor and stator.

However, the power electronic converter is limited to only 30% of the generator power rating. When the change in stator voltage is large, the rotor circuit can not follow the change in stator voltage. Unfortunately the power electronics make up a large proportion of the wind turbine capita therefore increasing the size of the power electronic rating is not a feasible option from an economic point of view and the rotor side converter must be protected from the high current transients. As seen in the simulation results, the rotor current transients raise the DC voltage. The generic DFIG model does not provide more detailed rotor current analysis but analysis of the DC bus link is proved to be sufficient in this respect.

5.3. Fault Ride-Through Protection

The power electronic converters require protection from the transient rotor currents that raise the DC bus link voltage of the power electronic converters. From foregoing analyses, during single phase-ground faults the DFIG can ride through the fault without additional protection, although the oscillations in DC bus link voltage may be of particular concern as the oscillations that occur on the rotor shaft may cause mechanical stress. During 3 phase-ground faults, protection of the DC bus link capacitor and IGBT’s are inevitable since the rise in the DC voltage is more significant. For this purpose a control scheme can be adopted to achieve this after analysing the behaviours of the detailed model of the wind farm during faults which were illustrated previously and which took the voltage at the limit of the fault ride through profiles.

As mentioned previously, the 3 phase symmetrical fault represents the most severe type of faults. In this paper, therefore, only 3 phase symmetrical fault on the 20 kV bus bar is investigated with power electronic protection scheme implemented. The scheme adopted is known as a crow-bar circuit [9]. In essence, a crowbar circuit is operated by adding a short circuit or low resistance path across the voltage source and is implemented by disconnecting the rotor side converter and short circuiting the rotor windings. The rotor side converter can be disconnected via a 3 phase contactor in the simulation model and the rotor windings can also be short circuited via a trip circuit. Simulation results with the protection circuit are shown in Figure 8 for comparison of that without such a scheme as shown in Figure 5.
With the fault ride through technique implemented, the DC bus link voltages is protected from the rotor over currents that rise the DC bus voltage during 3 phase to ground faults. Therefore damages to the IGBT’s is avoided. However, when the rotor-side converter is re-engaged after fault clearance, there is still a slight rise in DC bus voltage which may potentially rise the voltage to that without the protection scheme.

6. SUMMARY

A study of the DFIG’s behaviours responding to grid disturbances has been described in this paper. Grid disturbances studied include voltage sags from symmetrical and unsymmetrical faults. Simulations results confirm the DFIG can ride through the faults defined in the UK grid codes with the help of a crowbar protection circuit.

REFERENCES


Predictive Simulation for Model Based Energy Consumption Optimisation in Manufacturing System and Machine Control

Anton Dietmair*, Alexander Verl, and Philipp Eberspaecher
Institute for Control Engineering of Machine Tools and Manufacturing Units (ISW)
Universität Stuttgart
70174 Stuttgart, Germany

ABSTRACT

Resource efficiency and most of all energy efficiency have recently come to the focus of attention of production engineers, especially from the perspective of lean methods. As a prerequisite for optimisation, we show how the energy consumption of various production machines can be accurately predicted and how single machine energy consumption models can be integrated into models of more complex manufacturing systems with multiple machines and multiple types of energy carriers like electricity, pressured air and coolant. We go on to introduce model based energy consumption optimisation strategies that take into account goals in time, quality and cost and provide for an absolute reduction of the amount of energy consumed as well as peak power flattening. To maximise the operational flexibility while optimising the energy consumption, we focus not on permanent measures, but on methods that optimise the controllable machine and plant parameters in a situation-oriented manner. We conclude our paper with guidelines on how the techniques can be applied in industry.

1. INTRODUCTION

Resource efficiency and most of all energy efficiency have recently come to the focus of attention of production engineers and enterprises worldwide. Partly, this is due to the legislative pressure to raise energy efficiency [1]. But at the same time, innovative providers of high quality manufacturing solutions, see economic and ecologic sustainability no longer as antagonistic goals but as an opportunity to compete based on total life cycle cost instead of investment cost [2]. At the same time, lean principles have always been a core tool to achieve efficiency in manufacturing and beyond by identifying and eliminating waste [3]. To reach its full potential for continuous optimisation of production processes, lean production today relies on prediction and simulation based techniques [4]. For the particular case of sustainability in manufacturing, the energy consumption of manufacturing machines and systems has to be predicted and methods have to be provided to minimise the amount of energy consumed.

Therefore, based on the energy consumption model we presented during the FAIM2008 conference [5], in this paper we demonstrate how the energy consumption of production machines can be accurately predicted and how these models can be integrated into models of more complex manufacturing systems with multiple machines and multiple types of energy carriers like electricity, pressured air and coolant. Predictive simulation together with sensitivity analysis provides the basis for the accurate prediction of manufacturing system energy consumption that is required for energy procurement. We go on to introduce model based energy consumption optimisation strategies that take into account other goals like processing time, workpiece quality and machine and tool wear and provide for a reduction of the amount of energy consumed as well as peak power flattening. To maximise the operational flexibility while optimising the energy consumption, we focus not on permanent measures, but on methods that optimise the controllable machine and plant parameters in a situation-oriented manner. We conclude our paper with guidelines on how the techniques can be applied, either by manufacturing enterprises or by energy efficiency service providers.

* Corresponding author: Tel.: (0049) 711-685-84530; Fax: (0049) 711-685-82808; E-mail: anton.dietmair@isw.uni-stuttgart.de
2. ENERGY CONSUMPTION MODELLING AND PREDICTION

2.1. ACCURATE MACHINE TOOL ELECTRICAL POWER CONSUMPTION MODEL

The electrical power consumption of a production machine is given as the sum of the consumption of individual components of the machine. This, in turn, is determined by the operational state, which defines which components are active and thus consuming a certain amount of basic power, as well as the movements and manufacturing process executed by the machine. Figure 1 shows the corresponding elements of an energy consumption model and the interdependencies present between these elements.

As has been shown in [5], the basic consumption per operational state contributes a substantial proportion to the total energy consumed by a typical production machine. It can be modelled based on a state-transition graph and activity flags or power intake levels per component and state (Figure 2). This is typical for production machines today, which often have a large offset power intake independent of the manufacturing process [6]. Such a discrete event model without any extensions can already be used to optimise the operation of machines and production plants and to determine the relevance of the individual components for the total energy consumption [7]. An important aspect of the discrete event modelling consists of including the relevant time behaviour, e.g. the minimum time span a machine stays in a run-up state before being able to enter the next state, or stochastic mean times before transition.

Moving parts of the machine and the work piece contributes a second major component of the total energy consumption of a production machine. The motions and inertia forces themselves can normally be approximated using only kinematic modelling. In cases of highly elastic components, using simplified dynamical modelling is appropriate. Modelling and predicting the motion induced power consumption is of special interest for quantifying the energy savings that can be realised through energy recovery systems that feed back power into the grid during deceleration as well as light weight engineering methods that seek to minimise the amount of moving mass [8].

Figure 1: The electrical power consumption of a machine depends on its operational state, motions and processes.

Figure 2: Basic electrical power consumption per operational state and component for a milling machine.
Equation 1 can be used to compute the instantaneous electrical power consumption based on moving mass, acceleration, velocity and friction, and Figure 3 shows the good correspondence for the example of a milling machine:

$$P_{el,\text{motion}}(a, v) = \eta_{el} \left[ P_{\text{mech,acc}}(a, v) + P_{\text{mech,fric}}(v) \right] = \eta_{el} \left( m \cdot a \cdot v + F_{\text{fric}} \cdot \max\left(\left| v \right| \cdot \left| \text{sgn}(v) \cdot v_{\min} \right| \right) \right)$$  \hspace{1cm} (1)

where $P_{el,\text{motion}}$ = electrical power consumption due to motion, $a$ = acceleration, $v$ = velocity, $\eta_{el}$ = degree of efficiency of the motor and drive, $P_{\text{mech,acc}}$ = mechanical power required for acceleration, $P_{\text{mech,fric}}$ = mechanical power required to overcome friction, $m$ = accelerated mass, $F_{\text{fric}}$ = friction force and $v_{\min}$ = low speed model boundary.

While all shares of the total power intake introduced so far are independent of the benefit of the operation of a machine, the energy for the manufacturing processes not only contributes to the total energy consumption, but also quantifies the benefit of the machine. The nature of the manufacturing process and the equations for the computation of the processing power depends on the machine and its application. While in the case of a handling device the power equations will be similar to Equation 1, specialised equations can be given for processes modifying the workpiece by forming, cutting, grinding or milling, and even non-conventional tools like lasers can be modelled in their energy consumption [9]. A few publications on the influence of process parameters on the total energy efficiency of the manufacturing system exist [10]. In cases where no special results on the energy efficiency of a manufacturing process has been published, sufficiently accurate predictions about the power consumption can often be made based on equations for process planning and coefficients from standard handbooks, e.g. [11], as the example for the cutting power of a milling machine in Figure 4 and Equation 2 illustrates:

$$P_{el,\text{milling}} \left( d_{\text{cut}}, v_f \right) = \eta_{el} \cdot P_{\text{mech,milling}} \left( d_{\text{cut}}, v_f \right) = \eta_{el} \cdot d_{\text{cut}} \cdot v_f \cdot \frac{z_e \cdot k_c \cdot D}{z_w \cdot (\phi/360^\circ)}$$  \hspace{1cm} (2)

where $P_{el,\text{milling}}$ = electrical power consumption due to milling, $d_{\text{cut}}$ = depth of cut, $v_f$ = feed rate, $\eta_{el}$ = degree of efficiency of the motor and drive, $P_{\text{mech,milling}}$ = mechanical power required for the milling process, $z_e$ = number of cutting edges active in parallel, $z_w$ = number of cutting edges in total, $k_c$ = specific cutting force, $D$ = tool diameter and $\phi$ = cutting zone angle.

A number of other, more specific contributions to the overall energy consumption of production machines exist. One example are supply units that receive their electrical power through the machine and provide services like lubrication, but are not constantly running or linked to the operational states of the machine itself. In some cases, the temporal behaviour can be modelled based on the service provided and the usage of the machine. In most cases, though, their contribution is small compared to the basic consumption, motion and processing power.
2.2. Including Different Energy Carriers in the Model

A common mistake in energy consumption assessment of manufacturing systems is to limit the evaluation on the electrical power intake of an individual machine and to ignore non-electrical flows across the chosen system boundary as well as services that are necessary for the operation of the machine. It is also typical to study the degree of efficiency of supply units and to select particular solutions based on only the maximum possible requirements, while many supply units are actually operating at inefficient partial load conditions most of the time [12].

Typical non-electrical supply demands of production machines include pressured air, hydraulic oil pressure, lubrication, high and low pressure coolant and process heat. Some of these are put into the manufacturing system context for an example situation in Figure 5. It can be seen that some demands are covered by local supply units installed in or near the machines (typically hydraulics and lubrication, Supply Unit 1 and 4), while others are met by centralised units and factory-wide transport networks (e.g. pressured air, Supply Unit 5).
Central supply systems are often combined with local units for final conditioning of the supplied media. For example, this is typically the case for coolant, which is prepared centrally and fed to a large number of machines through a supply network. It is then pressurized by a local pump and fed into a single machine or a group of machines (Supply Units 2 and 3). After the used coolant leaves the machine, special devices separate chips from the fluid and feed the coolant back through a network into a recycling system. All mentioned steps require energy, and other forms of disposal and waste removal, e.g. vacuum devices for dust removal or structural cooling fluid networks, follow the same basic principles. In the case of structural or electrical cabinet cooling, it depends on whether the waste heat can be used as process heat or heating for buildings whether the heat is regarded as a loss. Services like workpiece handling, tool renewal and maintenance are required for machines to fulfil their tasks, so the energy consumed creating these services has to be taken included in the overall energy balance of the production. From what has been said it is obvious that judging the overall energy efficiency of a production system requires taking its nature as a network into account and quantify energy flows represented by different media and services.

Based on the structure of the power consumption model in Figure 1, this is easily possible. There, the total power intake consists of the individual power demands of the active components. These components can be either local, like electrical motors and drives moving the axes of a machine, or central supply units shared by machines over a network. In the former case, their electrical power consumption goes directly into the mains power intake of the machine itself, while in the latter case a proportion of the power consumed by the supply unit can be assigned to the machine indirectly via calculation of the supply demand of all connected machines. The role of component models in the energy consumption model is the transformation of a quantified demand of e.g. coolant with specific properties, e.g. pressure, and calculate the electrical energy required by the component to meet the demand. Internally, the structure of a supply unit model is similar to that of a machine and thus follows Figure 1 with manufacturing process modules replaced by demand fulfilment modules. Some supply units show a constant electrical power consumption for an operational state, while the power intake of others will depend on the flow of media drawn by the machines.

In most cases today, the control of supply units is strictly local and based only on the observed flow of media, pressure drops and the like, but the energy efficiency of supply networks can be raised if the demand is predicted and communicated by the machines. While fully centralised supervisory control for media scheduling contradicts the lean principles and would lead to reduced reliability, local communication of demands and robust supply control strategies are lean compliant and can contribute to waste elimination. This is illustrated by dashed lines in Figure 5.

2.3. COMPOSITE MODELS OF MACHINES AND MANUFACTURING SYSTEMS

In the previous paragraph, it has been shown that the energy consumption behaviour of a machine is determined not only by the machine itself, but also by the supply units. At the same time, the operation possibilities of a machine at any time are restricted by the behaviour and operational states of the supply units. Using the product graph of the individual operational state machines as a model for the overall behaviour is limited to systems with very few subsystems that each have only a very small number of states, as the number of states of the product graph will be very large otherwise. Using the graphical formalism that is well established for Petri Nets, with circles representing states, bars representing transitions and markers denoting the active states, Figure 6 illustrates possible approaches to building composite models from multiple subsystems. Following the direction defined by the graph, a transition is unlocked when all preceding states carry markers. As soon as the transition is activated by an external event, the markers from the preceding states are removed and new markers are added to the subsequent states. We apply this to link elementary models which each have one active state of their own, but share synchronous transitions [5].

![Figure 6: Composite models of functionally linked components (dashed) created with linked (a) or independent states (b).](image-url)
Figure 6a shows a simple example for a machine and its coolant supply unit linked by the transitions representing the events “switch coolant on” and “coolant switched off”. Both the machine and the supply unit now can have an energy consumption profile attached to their operating states, with the total power intake resulting as the sum of both. In addition to that, each operational state may have a more complex parameter dependent behaviour. For example, the energy consumption of the coolant supply unit may be dependent on the flow rate of coolant drawn by the machine (dashed line in Figure 6a). Sub-models may also be linked by functional dependencies, even if there are no control mechanisms linking their operational states. Figure 6b shows this for the example of a milling machine that receives its coolant from an independently acting supply unit, with the presence or absence of the coolant flow clearly determining the possible manufacturing processes and process parameters.

It has to be stated that interdependencies in manufacturing systems are not limited to the supply network, but are often also given by the workflow and by machines, handling units and other components interacting. These structures can also be modelled as a composite of a number of elementary operational state graphs. Figure 7 illustrates this for the example of a milling machine fed by a handling robot from two conveyor belts. Each operational state is assigned an energy consumption profile like the one introduced in Figure 2 and the total instantaneous power intake is given by the sum of the individual power consumption associated with all states carrying a marker. In most cases, synchronous transitions are sufficient to link the elementary models, like for the conveyor belts and handling robot in the example. More complex interactions, e.g. between the handling robot and the milling machine, can be defined using additional states. In the example given, this allows the robot to load a workpiece into the machine even if the machine is not yet ready to start milling. Additional states typically do not add to the total energy consumption.

3. SIMULATION BASED MULTI-OBJECTIVE ENERGY CONSUMPTION OPTIMISATION

Avoiding and eliminating waste is a core principle of lean production. Over-engineering and centralised control with the aim to raise efficiency would not only contradict the lean principles, but often lead to diminished production system reliability and productivity. Robust decentralised control solutions complemented by global system modelling and simulative optimization, in contrast, are viable strategies to detect inefficiencies and avoid waste. Together with [5], chapter two of this paper provides the framework for modelling the energy consumption behaviour of machines and manufacturing systems in a way that allows to optimise the usage profiles and control strategies. Optimisation must not only take into account the total energy consumption and peak power relevant for energy procurement and saving a large proportion of the total lifecycle cost of production systems [13], but also other objectives like machining time, workpiece quality, tool and machine wear, and other elements of the life cycle cost like investment. Some of these objectives are straightforward to assess in the context of the energy consumption modelling approach presented here, e.g. machining time and investment cost, while others like workpiece quality and wear require models of their own. Models in these areas are increasingly available, and depending on the manufacturing task, the information available about the production system and the goals, a tailored optimisation strategy can be chosen. While [5] named a number of possibilities for simulation model based energy efficiency optimisation in planning, in this section we will give examples on how model based automation can be used to raise energy efficiency.
3.1. SELECTING OPTIMAL MACHINE PARAMETERS

The first possible way to optimise the usage of machines is to determine machine and process parameters that satisfy the criteria in an optimal way. A typical machine parameters that can be subjected to optimization is the maximum acceleration for interpolation of numeric control programmes. Reducing the maximum acceleration leads to substantially lower peak power and reduced machine wear, while the machining time increases. The longer machining time means that the basic power of constantly running machine components and supply units is consumed over a longer time, eventually leading to higher total energy consumption per manufactured part. A process parameter that can be used to influence the amount of energy required per manufactured part is the feedrate. While a high feedrate leads to a short machining time and thus to a reduced duration of basic power consumption, going beyond a limit specific to the process, both the instantaneous efficiency and the quality of the machined surfaces is substantially diminished. Finding the strictly optimal parameter settings requires a very deep understanding of the manufacturing process and is not always practical [10], but a basic energy consumption model can contribute to improvement.

3.2. LOGIC AND SEQUENTIAL CONTROL PROGRAM OPTIMISATION

A second and more substantial possibility for energy consumption reduction lies in the programming of the logic and sequential control of production machines and systems. This can be illustrated by the example of an automatic hibernation mode of a production machine given in Figure 8. There, it is known at a specific point of time that the machine will be idle for a known time span. Two possibilities for energy saving action exist, namely to switch off the machine and then re-start it again (Figure 8a) or to switch off only one component that can be re-started instantaneously (Figure 8b). Switching off the machine in this example actually would increase the total energy consumption, as the time the machine is consuming no energy is short and a number of operational states with high power consumption have to be entered during the re-start. In addition to that, the machine would be available only after the scheduled time. Consequentially, an automatic control algorithm based on the energy consumption model would choose option b. More concepts for optimisation techniques that can be implemented in programmable logic control are given in [14]. In addition to machine controls, the sequential control of a manufacturing system including scheduling can be optimised using model based total energy consumption prediction as well [15]. During any automatic optimisation, aspects like usability and thermal stability have to be taken into account.

3.3. OPTIMISATION OF NUMERICAL CONTROL PROGRAMS

In the stage preceding execution of machining programs on machines and manufacturing systems with or without energy optimal logic and sequential control, Computer Aided Process Planning and Manufacturing (CAP/CAM) tools can be used to find the optimal strategy producing parts on a manufacturing system. As illustrated by Figure 9, Numerical Control (NC) simulation can be combined with the model presented in this paper to determine the energy consumption for any machining strategy and process alternative, e.g. dry machining. Optimization algorithms and software can then be used to find the variant with minimal energy consumption that satisfies all other constraints.

3.4. GUIDELINES FOR MODEL BASED ENERGY OPTIMISATION

The examples given above highlight the high potential that is hidden in model based optimization and automation to achieve substantial improvements in energy efficiency of machines and manufacturing systems in a lean way. The central question that has to be addressed is how machines and manufacturing plants can be operated with high efficiency under partial load conditions, as partial load conditions take up most of their active time [14].
Selecting the optimal level of detail for practical model-based optimization can be achieved by following lean principles in applying the framework introduced here. Different approaches can then be followed to optimize the usage and control of manufacturing systems. These include sensitivity analysis to determine the parameters that have a big influence on energy consumption and other goals, pareto analysis, search heuristics and evolution strategies to find the attainable global optima and breaking down the problem into robust operational state-wise optimization.

4. SUMMARY

In this paper, we have introduced the framework for accurate modelling of production system energy consumption, with special emphasis on the dynamic elements of manufacturing machine power and the inclusion of non-electrical energy. We have gone on to demonstrate how composite models can be built out of simple elementary energy consumption models of individual machines and components. Based on examples for the simulative optimization of controllable parameters, we have given guidelines for model-based energy consumption optimization.

REFERENCES

Bionics – An Inspiration for Intelligent Manufacturing and Engineering

Bill Gaughran and Sonya Quinn

Department of Manufacturing & Engineering
University of Limerick
Ireland

ABSTRACT

Value creation in all its facets lies at the core of intelligent manufacturing and engineering. In the last twenty years the field of manufacturing has undergone many changes and refinements. Terms such as Integrated Management Systems (IMS), Just in Time (JIT), Toyota Production System (TPS) in the context of Lean Production and ‘Flow’ were parts of the toolset developed by the Toyota Corporation which pushed them to the forefront of world automotive production. While benchmarking the design production systems and their associated efficiencies is very worthwhile, there are other engineering design, lean production, just in time, and production and supply chain exemplars which are worth investigating. A primary source of best-practice engineering in flexible and intelligent manufacturing is to be found in the study of ‘Bionics’ (Biomimicry). The intelligence in design and operational efficiency which is brought to Bionics by design in nature was recognised by Leonardo DaVinci when he wrote: “…in her (design) nothing is lacking and nothing is superfluous” [1]…This paper examines how design and engineering can learn and apply through the study of bionics/biomimicry, a vast pool of knowledge of design and systems engineering strategies. Such strategies and exemplars will provide benchmarks which will result in inspirational approaches in design, efficiency and sustainable engineering solutions.

Keywords: Efficiency, Sustainability, Energy Management, JIT, Lean, Bionics.

1. INTRODUCTION

Long before contemporary terms such as bionics, biomimicry, biomimetics, biognosis or bionical creativity engineering became well-know, people have appreciated the beauty and design in all aspects of nature. What is Bionics or Biomimicry? (From bios – life, and mimesis – to imitate). It may be described as an innovative approach to design and engineering, inspired by design in nature. The wise King Solomon of ancient Israel stated more than three thousand years ago that: “…there is nothing new under the sun” (Eccl. 1:9). It might be claimed that this ancient had little to compare with, and the wonders of modern technology were long in the future. However, today and into the future engineers and designers will do well to pay strict attention to specific detail and systems design in nature so as to draw inspiration for efficient, renewable, self sustaining, and ultimate solutions to problems of environmental sustainability. ‘Innovation inspired by nature’ [2], can lead to a new path of discovery and problem solutions, which for the most part have been overlooked. Nature is the ultimate exemplar in ‘eco-design’, and deserves the full attention of all who endeavour to tackle the problems of sustainable technologies in energy, materials, medical engineering and technological innovation. Why? Well when we examine closely its designs and systems we see that nature:

- runs on sunlight;
- uses only the energy it needs;
- fits form to function
- recycles everything
- rewards cooperation
- banks on diversity
- demands local expertise
- curbs excesses from within
- taps the power of limits

A design specification such as the above would frighten any team of experts, and yet this is what we find in nature. When people appeared to lead more leisurely lives, they seemed to appreciate these facts more
than we do today. Even in rudimentary ways DaVinci (1452-1519) tried to mimic flight (see Figure 1). It didn’t fly, but he was thinking in the right direction.

One of the good things about modern technology is that it does provide greater opportunity to mimic natural design and to test its effectiveness. Towards that end this paper will look at natural and applied examples from the skies (birds) the seas (fish/mammals) and the land. All of these, make up the entire eco system, but each has its own ecosystem and in turn sub-systems.

Another ancient character who recognised this was the man Job, known for his patience, Moses wrote in the book of Job, nearly five hundred years before King Solomon:

“Ask please the domestic animals and they will instruct you; also the winged creatures of the heavens, and they will tell you. Or show your concern for the earth and it will instruct you; and the fishes of the sea will declare it to you.” Job 12:7-8.

In examining some of the examples in nature that Job observed, and in the context of available, and yet to be developed contemporary technologies, it can be seen how these may be harnessed to inspire new approaches to design and engineering. Approaches, which in the context of mimicking design in nature, have the potential to fulfil some or all the criteria listed in paragraph one above, and so provide engineering solutions which are energy and environmentally sustainable.

2. THE WINGED CREATURES AND THE FISHES

While DaVinci’s ‘flying machine’ never got to fly, birds/mammals such as the seagull and the bat have been the inspiration for advancing efficient technologies. Funded by the U.S. Air Force and NASA, US aerospace engineers have built prototypes of 150 mm to 600 mm drones capable of squeezing in and out of tight spots in cities — like tiny urban stunt planes. Their secret: seagull-inspired wings that “morph,” or change shape, dramatically during flight, transforming the planes’ stability and agility at the touch of a button on the operator’s remote control [3].

Currently at the University of Limerick the ‘cranked’ wing is being tested on large model (1800 mm) aircraft, to determine the manoeuvrability and fuel consumption. Additionally the model is being tested with and without leading edge wing-‘tubercles’, this idea is inspired by the very large but highly manoeuvrable Humpback whale. Wind tunnel tests of model humpback flippers with and without leading-edge tubercles have demonstrated the fluid dynamic improvements tubercles make, such as a staggering 32% reduction in drag, 8% improvement in lift, and a 40% increase in angle of attack over smooth flippers before stalling. A company called WhalePower is applying these lessons to the design of
wind turbines and fans of all sorts – industrial ceiling fans and other HVAC systems, computer fans, etc. – to improve their efficiency, safety, and cost-effectiveness [4] In Australia the use of fan blades with ‘tubercles’ used in air conditioning systems, save significantly on energy and run much quieter. These giant creatures of the seas may weigh thirty-six tones, and measure up to fifteen meters in length. Figure 3 below shows an illustration of the ‘bubble net’ of the whale (d), close-up of his flipper (a) and a line diagram of the hydrodynamic flow pattern comparisons between a smooth leading edge and a tubercled one (b). Also shown is a close up of a wind turbine blade (c) for electricity generation with a similar leading edge. In addition to increased efficiency, the latter also contributes to overcoming the general objection of noise pollution from conventional turbine blades.

The airplane experiment described above will attempt to reduce drag and so save fuel and thus CO2 output. The combination of the cranked and tubercled wing, appear to have good potential, the tests are ongoing. If the hypotheses are correct there are likely to be highly significant gains in aircraft technology and efficiency.

We have also much to learn from birds and sea creatures about how to use and transfer energy. Returning to the seagull the question arises: why this and other birds who may be constantly standing on ice do not freeze, or why some fish in extremely cold seas continue to function well. The answer lies in their built in counter-current heat exchange design. The veins and arteries pass in contact or may be linked by shunts, where the warmer blood in the arteries heats the cold in the veins. For example by the time the cooler blood reenters the body of the gull from the ‘cold feet’, it is only two degrees Celsius cooler. Figure 4 shows diagrammatically the heat exchange mechanisms in gulls and fish.

The applications of heat exchange are numerous and highly beneficial in the efficient running of low energy heating systems. When, for example, a geo-thermal or air-thermal heating system is installed, and the heat pump unit has elevated the ground temperature from 12° Celsius to 22° degrees Celsius, it would be extremely wasteful if the exhaust air or fluid were simply returned to source without having a system of elevating the temperature of the incoming replacement. This is accomplished by the seagulls legs counter-flow heat exchange principle.

This is just another example of the benefits of looking at design in nature through the observant eyes of men such as Job or Leonardo DaVinci.
The applications resulting from such understanding, are used in many heating systems - shown in Figure 5 below are diagrams of a geothermal energy system for a home and the heat exchange unit which adds to efficiency by significantly reducing lost energy.

Mammals, both sea and airborne offer an infinite source of inspiration to engineers and designers. This is already being applied in sonar and radar systems which keep us safe while ‘flying’ or assist trawlers to locate fish etc. Hikers moving through unfamiliar territory are well advised to carry a compass, if not a GPS unit, to stay on course. Other animals appear to have complex navigational equipment in-built. New research reveals that Israel's blind mole rat (Spalax ehrenbergi) uses the Earth's magnetic field on long journeys, much like a compass, to continuously monitor and maintain its course. But that's not where the burrowing rodent's abilities end. The mole rat also has an uncanny habit of burrowing around obstacles—such as ditches or concrete blocks—without ever coming in to physical contact with them. Studying how this is accomplished is very important for human tunnelers and miners. Making navigational mistakes underground is expensive and time consuming, and of course energy consuming. It is estimated that excavating soil uses between 360 and 3,400 times as much energy as moving the same distance above ground [11]

Bats, porpoises and dolphins have biological sonar technology that is vastly superior to anything that man has yet devised. They use sonar to distinguish resolution and identify objects and their material characteristics for objects that are, to all extents and purposes, bewilderingly similar. Researchers in the University of Leicester in the Department of Geology are involved in a project to develop sonar systems based on those animals.

Researchers recording echolocation calls from bats in free flight hope that such technologies will have far-reaching effects in medical and geological imaging, focussing on materials characterisation and non-destructive evaluation. This will particularly apply to the search for oil, gas and other natural resources in rock samples, and may have a role in the waste management industry. Their funding from the Engineering and Physical Sciences Research Council (EPSRC) is intended to apply acoustic
capabilities that occur in the natural world to engineering concepts, such as biomedicine and underwater sonar [12].

It appears that dolphins are able to form very high definition ‘images’ through their sonar feedback and perhaps bats can do likewise. Can this technology be carried further in an application to assist unsighted humans? Perhaps the blind crewman (below Fig. 6) in StarTrek may not be as far fetched a piece of science fiction as once thought. ‘Primitive’ systems are already under development and trial, and perhaps in the not-too-distant future a high definition product may be developed.

![Image of dolphin and sonar feedback](image1.png)

**Figure 6** Can the high definition sonar images of dolphins be applied to help blind people ‘see’? [13][14]

In an important breakthrough in deciphering dolphin language, researchers in Great Britain and the United States have imaged the first high definition imprints that dolphin sounds make in water. John Stuart Reid, and Florida-based dolphin researcher, Jack Kassewitz, has been able to image, for the first time, the imprint that a dolphin sound makes in water. The resulting “CymaGlyphs,” as they have been named, are reproducible patterns that are expected to form the basis of a lexicon of dolphin language, each pattern representing a dolphin ‘picture word’ [13].

3. **THE GECKO STORY**

The Gecko is a lizard whose feet soles are covered with modified layers known as lamellae, each of these is then covered with similarly oriented tiny hairs known as setae. Each setae splits into as many as 1000 branches with spatula shaped tips, measuring a mere 200 nanometers wide. These billions of setae tips create such a vast surface area and are able to nestle so closely with their target surfaces, that the many weak interactions add up to a significant adhesive force.

![Image of Gecko foot](image2.png)

**Figure 7** Gecko Foot inspiring robotic advancement a [15] b, c [16] d [17]
Their grip is so strong that "scientists calculate that a gecko's-worth of setae, which would fit on a small coin could lift about 115 kilograms" (Ben-Ari, 2002). Another way to try and picture it: the combined charge is so powerful that, theoretically you could suspend a 40 kilogram weight from the gecko and they could still get across the ceiling (Robbins, 2001). Therefore in trying to duplicate this wonder scientists concentrate on "the smaller the hairs are, and the more of them you have, the greater the adhesion." (Ron Fearing, engineer at the University of California, Berkeley, as quoted in McDonagh, 2003). [18]

Scientists have discovered that an appropriate arrangement of setae and spatulae held the geckos to the wall by means of a type of a intermolecular attraction known as a van der Waals force. The same force that holds geckos to walls has been used to explain everything from snowflake formation to spider acrobatics.

A bandage based on the gecko grip principal has been developed. Nanoscale features molded into the surface of the adhesive bandage help it bond extremely well. Similar features on the feet of geckos allow the lizards to walk on walls and ceilings. The phenomenon of self-cleaning in gecko setae is surprising because setae are adhesive but can self-clean when dry. Adhesion in gecko setae is a consequence of many divided contact points (spatulae) that deform to achieve intimate, high-density contact with the surface, whereas lotus-like surfaces remain slippery because their rough, and in some cases waxy, cuticle prevents intimate contact. Lotus-like surfaces require water as a cleaning agent, whereas self-cleaning in gecko setae may occur because it is energetically favorable for particles to be deposited on the surface rather than remain adhered to the spatulae. [20]. It is of interest to note that both gecko and lotus-like surfaces become cleaner after contamination for structural reasons; both possess a micro- or nano-rough topology that reduces adhesion with solid and liquid surfaces alike. [21]

There's no gluey stickiness to the gecko's foot, nor does it use suckers or some strange muscular ability - you can detach the part of the foot that gives it the ability to walk up plate glass (harmlessly to the creature - it re-grows like hair) and use it to make "gecko tape" that sticks to nearly anything - instead it relies on a tiny force at the nano-level, multiplied up by the use of a vast number of tiny surface contacts [17]

4. BEES AND OTHER THINGS

Honeybees craft the marvelous structure of the honeycomb (Figure 10) to very high tolerances and specifications. Each cell has two vertical walls, with "floors" and "ceilings" composed of two angled walls. The cells slope slightly upwards, typically about 13 degrees, towards the open ends. There are two possible explanations for the reason that honeycomb is composed of hexagons, rather than any other shape. One, given by Jan Brożek, is that the hexagon tiles the plane with minimal surface area. Thus a hexagonal structure uses the least material to create a lattice of cells within a given volume – maximum volume for minimum materials. Another, given by D’Arcy Wentworth Thompson, is that the shape simply results from the process of individual bees putting cells together: somewhat analogous to the boundary shapes created in a field of soap bubbles. [22]

The closed ends of the honeycomb cells are also an example of geometric efficiency, albeit three-dimensional and little-noticed. The ends are trihedral (i.e., composed of three planes) pyramidal in shape, with the dihedral angles of all adjacent surfaces measuring 120°, the angle that minimizes surface area for a given volume. Tolerances are amazing when considering the dimensions – the worker cell is 5.2 mm across the flats and the drone cell 6.2 mm across, with a cell
wall thickness of 0.073 mm and 0.002 mm tolerance on wall thickness. The honeycomb is indeed a structure of amazing precision and demonstrates the ultimate economy in use of materials and conforms to an arrangement of minimum potential energy. The strength weight ratio is an example to engineers in lightweight wall structure, as in the construction of structural panels in aircraft. See figure 11a. Shown in Fig. 11c is a soap bubble structure between two plates of glass, note the hexagonal structure (minimal energy). 11b would, as a section of dried cracked earth appear chaotic, however the 120° angles are clearly visible, so even in the apparent chaos of cracked earth, nature minimises the energy expended.

5. Fibonacci, Close Packing and The Golden Section

When the Italian mathematician Fibonacci discovered that the number series for which he is now famous, the implications of this in relation to design were poorly understood. Even now this mathematical number series is to be found again and again in design. It produces a close packing solution, a proportion such as that found in the nautilus shell, the sunflower, the pattern which the leaves grow in order to expose them to sunlight, or the ever pleasing aesthetic of the golden rectangle as seen in Figure 12.

This proportion has been used aesthetically in art and architecture from the Greek Empire right to this day. It is an inspiration for engineering design problems such as less noisy and significantly more efficient electric fans. In the nautilus shell it is not just the Fibonacci curve which in itself is beautiful, but each chamber appears to increase in volume in a 1:1.618 ratio. There is indeed an infinite source of exemplars to be found in nature.

6. CONCLUSION

As we strive to combat global warming, develop sustainable energy systems, protect the environment and Earth’s eco-systems, we have at our disposal in the natural environment all the exemplars we need.
Companies are beginning to see the possibilities as they look to nature for design inspiration. Recently the Mercedes Benz motor company built an aerodynamically efficient family car based on the hydro-efficient boxfish, see Fig. 13 below. This paper has only discussed a few of these designs, and not even in minute detail. It behooves engineers and designers to look closer at such a rich pool of inspiration.

When DaVinci said that “in her inventions nothing is lacking and nothing is superfluous” he was accurately describing design in nature.

Now, innovation has never been more necessary, so whether we look to the great rainforests or the humpback whale, or the recently discovered flagella motor, which is just 20 nm across, (a fraction of the diameter of a human hair). This complete motor, which is described as a machine of ‘irreducible complexity’, has serious potential in nano-medical treatment and investigation. This “outboard motor”, which is used by bacteria to propel through liquid, as its propeller rotates at 400 times per second, gives us some insight as to what is possible when nature is observed under the microscope. It can be said with certainty that nature is lean, just-in-time, the ultimate problem-solver, and the perfect exemplar in sustainable engineering, and is indeed an inspiration for intelligent and innovative engineering.

References

1. Da Vinci, Leonardo (1452-1519)
17. http://www5.pbrc.hawaii.edu/microangelo/gecktoes.htm
18. Ron Fearing Engineer, University of California, Berkley as quoted in McDonagh 2003
Environmental Management Systems in Smaller Enterprises

(Problems and Solutions)

Mike Gallagher, William Gaughran and Pat Phealan

Department of Manufacturing & Engineering,
University of Limerick
Ireland

ABSTRACT

On average SMEs represent over 90% of all enterprises and account for 50-60% of employment at an Irish level. SMEs are particularly important in supporting economic growth and livelihoods in countries. Small enterprises in particular form the back bone of the Irish economy, with eighty-two percent of all industrial enterprises in Ireland classified as ‘small.’ Of which, 95% are Irish owned employing between three and forty-nine people in 2004. [1] Due to their size, location and ownership SMEs have been seen to be embedded in their communities. This brings with it a level of acceptance by their stakeholders, and in particular a certain neglect toward their environmental performance. This coupled with the reality that SMEs are less likely to be subjected to rigorous inspections in relation to environmental standards, forms a type of comfort zone in which many wait and detracts from the real truth of the environmental situation. In various forms SMEs learn from larger companies, for example business management systems are tailored for the larger companies and do not fittingly suit SMEs. [2] This holds true for environmental management as unlike large companies, SMEs do not have the expertise or resources to implement a formal management system successfully. [3] Studies confirm the larger the company the more likely it is of having an EMS implemented [4] This paper will identify & discuss the barriers that face SMEs in implementing a formal EMS, in particular ISO 14001. The initial stages of research currently undertaken in developing a strategic framework for SMEs in achieving sustainable development are discussed. SMEs and their current impact on the environment is examined, supplemented by regional survey results focusing on a sector specific case studies addressing SMEs current awareness and practices. EMSs, their origin, structure, motives for adoption results and finally progress to identifying the precise barriers that exist for SMEs in achieving ISO 14001, is also discussed.

1. INTRODUCTION

Individually SMEs impose negligible stress on the environment however collectively their impact is considerable. Some researchers have suggested that SMEs could contribute up to 70% of all industrial pollution (although this is not irrefutably supported by evidence). [5] In their annual report ‘SMe-nvironment’ on the environmental performance of British SMEs - NetRegs state that “the smaller a business is, the lower its level of environmental awareness and the less likely it is to take action to address it’s environmental impacts.” [6] SMEs operate quite differently to one another when it comes to environmental practices. Palmer and Hillary (2000) classified SMEs in the manner in which they dealt with their environmental performance. [7]
These stages categorised SMEs environmental activity/inactivity. However, it still provided a rudimentary understanding of the actions of the SMEs. With the majority of SMEs classified as ‘inactive’ [8] along with the heterogeneous nature of SMEs it is difficult to draw any generalisations into:

1) practices or principles
2) environmental performance levels,

employed by SMEs within any of these stages. To assist SMEs in reaching the latter stages of ‘managed’ and ‘standards’ e.g. ISO 14001, it is imperative we identify firstly the barriers experienced, and secondly develop tools to overcome such grievances. SMEs represent an array of assorted sectors, environmental impacts that exist in one sector may not be the same in the other. In order to achieve the above objective we must hone in on the properties of a particular sector and then project them onto general findings. [9]

WOOD MANUFACTURING SECTOR [ENVIRONMENTAL PERFORMANCE]

As well as lagging behind their larger business counterparts in regard to their environmental performance wood manufacturing SMEs have also been seen to lag behind alternative sub-sector SMEs in surveys conducted in the UK. [6] Their poor environmental performance is supported from results received from an Irish regional study undertaken in late 2007. The purpose of the Irish investigation was to gather, interpret and ascertain regional and sector specific information regarding, Irish Timber Frame Manufacturing (TFM). The study established SMEs current knowledge and practices relating to the environmental issues. The majority of SMEs have a workforce of below fifty employees with the mean average calculated at 38 employees. This classes 68% of the study within the EU’s definition of a ‘small enterprise’ supporting the prevalence and importance of small enterprises. Fig. 2 illustrates the number of SMEs participating in listed environmental performance measures. There is not one measure, including recycling, in which they all participate. There is a severe lack of activity when it comes to TFM SMEs implementing simple measures to curb their negative environmental impacts (NEI). This is supported by 88% not responding to a question directed toward the NEI activities they perform. Either 88% of TFM SMEs do not commit any of these activities, though this is unlikely, or a very high percentage carry out these activities not wanting to admit it on paper the latter being a more realistic conclusion. The detrimental activities that cause NEI were as follows:

- Store chemicals oils or fuel
- Produce and or import packaging
- Use water pumped from lakes & rivers
- Emit smoke or fumes
- Discharge to water course or sewer
- Store waste

![Figure 2. Environmental Measures Undertaken by Irish TFM SMEs](image)
In relation to EMSs, 27% of companies within the Republic of Ireland had a type of EMS, in contrast to 67% in Northern Ireland. Suggesting that NI TFM’s have a deeper knowledge and understanding of EMSs and considering the degree of similarity in business to their Irish counterparts, a superior environmental performance. To breakdown the results of the different EMSs further:

- 0% had ISO 14001,
- 12% have BS-8555 EMS implemented,
- 20% have either an in-house EMS or an unspecified EMS,
- 28% would like to implement an EMS
- 40% do not want to implement any type of EMS.

96% of respondents requested to receive a summary of the survey signifying a high level of interest into environmental performances and practices of their competitors/colleagues. However, it does not take away from the fact that SMEs and in particular wood manufacturing SMEs, are an environmentally poor performing sector.

2. ENVIRONMENTAL MANAGEMENT SYSTEMS

To identify and fully comprehend the difficulties SMEs experience during implementation, a firm understanding of the EMS is essential. It is beneficial at this stage to explain the origin, structure, and actions experienced through the implementation of a formal EMS are examined in the context of this sector specific investigation.

An EMS is a set of management procedures to analyse, control, and reduce the environmental impact of a company. An EMS is said to be appropriate for all types of organizations of varying sizes. Starkey (1998), defines an EMS as consisting of a number of interrelated elements that function together to achieve the objective of effective environmental management. [3] Environmental management began in North America in the 1970’s in response to stringent and fast-growing environmental legislation. [10] A plethora of in-house EMS followed until the first attempt of standardization arose in 1992, when the British Standards Institution introduced BS 7750. This was the first environmental management standard ever, [10] it was drafted from the English equivalent of ISO 9001 and BS 5755 [11] BS 7750 was soon adopted as a national standard by various countries, e.g., Denmark and the Netherlands. From BS 7750 EMS various environmental management national standards were created, e.g., the French X30-200 and the Irish IS 31. [10]

In a growing mismatch of national EMSs in the early 1990’s, there was a need for a European or international standard. This was the origin of the two current formal EMSs - ISO14001 and EMAS. Companies now had the ability to choose from European and/or an International standard of environmental management and inevitably the national standards and in-house EMSs fell to the wayside - with the original EMS BS 7750 being withdrawn in March 1997. [10] EMAS certifications are much lower than ISO 14001 and this position is likely to remain with its popularity peaking in 2001. [12] Certifications of ISO14001, which became effective after EMAS, are now currently greater than 32:1. [10] Re-affirming its status as the world’s most recognized and prominent EMS.

**EMS ISO 14001 STRUCTURE**

The majority of EMSs is derived to some extent from the ISO 14001 model. The model is based on PDCA cycle first developed by deeming. The Plan –Do –Check –Act cycle’s repetition gives life to standards ability to achieve continual improvement. The model forms the structural framework to the standard the framework has five mandatory parts. [13]
Environmental Policy: Creation of an ethos for the organisation as a whole, specifying its commitment to continual environmental improvement.

Plan: establish the objectives and develop the processes necessary to deliver results in accordance with the organization's environmental policy.

Do: implement the processes.

Check: monitor and measure processes against environmental policy, objectives, targets, legal and other requirements, and report the results.

Act: take actions to continually improve performance level of the environmental management system. [3]

What has an EMS to offer and what are the main reasons for EMS adoption? It firstly enables a company to become compliant to all relevant environmental legislation. Building on this it offers the company increased efficiency and finally it reduces costs and provides for greater operational consistency. Table 1, sourced from work by Hallinan and Starkey, provides a comprehensive list of benefits experienced by SMEs. [3, 8]

<table>
<thead>
<tr>
<th>Cost Savings</th>
<th>Cost savings can be achieved through a number of ways such as introducing more efficient processes, and thus reducing waste or resource usage, such as raw materials or electricity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensuring legislative Compliance</td>
<td>By ensuring legislative compliance, the company can avoid possible fines for non compliance</td>
</tr>
<tr>
<td>Anticipating future Legislation</td>
<td>By anticipating future legislation a company can take time in taking investment decisions instead of a more costly last minute drastic change.</td>
</tr>
<tr>
<td>Reduced Environmental Risk</td>
<td>Undertaking principles by certain Environmental Management System (described later), environmental risk can be reduced possibly improving relations with investors who base their decisions on risk.</td>
</tr>
<tr>
<td>Meeting Supply chain requirements</td>
<td>More and more companies are demanding that their suppliers adhere to certain environmental standards, and if these are not met, the company will most likely lose their contract. Thus, environmental management can help SMEs stay in business, and gain new contracts.</td>
</tr>
<tr>
<td>Improved relations with regulators</td>
<td>When a company shows a good level of environmental management it is possible that the regulatory bodies take a ‘nicer’ approach to them e.g. a reduction in the number of inspections per year.</td>
</tr>
<tr>
<td>Improved Public Image</td>
<td>Environmental concerns are becoming more and more important for customers and hence a company can improve its image by publicising its environmental activity.</td>
</tr>
<tr>
<td>Increased Market Opportunities</td>
<td>A combination of the above factors such as reduced costs and improved public image can open new market opportunities.</td>
</tr>
<tr>
<td>Employee Enthusiasm</td>
<td>As the environment its becoming an increasing concern in recent times, employees are often very enthusiastic that they can contribute towards improving environmental performance</td>
</tr>
</tbody>
</table>

Table 1 EMS benefits/advantages (Source Cross 48, 76)
Currently Formal EMS adoption is below 0.5% of all European SMEs. [14] Figure 3 by Russomanno perfectly illustrates the motives behind, the barriers that present themselves and the results that are realised through the adoption of a formal EMS. [15] Motives for adoption may be categorised into two group’s intrinsic and extrinsic motives. Studies reveal the majority of SMEs are pressurised to adopt ISO 14001 through extrinsic motives such as stakeholders, supply chains pressure, etc. [5] This type of adoption is harmful as there is an unwillingness on behalf of the management and employees to implement the EMS and sometimes has been seen to hamper the effectiveness of the EMS. [9] Some of the benefits within the table e.g. employee enthusiasm and improved public image are regarded as intrinsic motives. It is interesting to note that some Irish TFM SMEs who extrinsically adopted an EMS realised intrinsic benefits through its implementation. This reflects a repeating theme in the research of environmental management that a severe lack of knowledge by the SMEs leads to fallacious assumptions of the standard and what it truly can achieve. [9] This is further more supported by Irish findings were companies without any type of EMS suggested that ISO 14001 offered extrinsic benefits only with some stating it provides no benefits at all. On the other hand companies with an EMS, and therefore hold a level of environmental management knowledge, provided a list of both intrinsic and extrinsic benefits.

To counteract this lack of knowledge many seek external assistance. [14] A TFM SME who received such assistance believes that without it achieving BS 8555 (a type of EMS) would have been unattainable for them similar SMEs. Figure 3 [8] and other studies discussed also identify SMEs lack the necessary financial requirements to achieve a similar environmental feet as their larger counterparts [3] On a comparison study of implementation barriers experienced by SMEs in Hong Kong, Canada and globally lack of ‘environmental knowledge’ and ‘financial cost’ all appear in each list of top barriers. [14] [16] [17] Understanding 80% of SMEs implementing ISO 14001 opt for external assistance this brings its own notable financial problem. [14] As many have difficulties in obtaining appropriate subsidised or free assistance. In Hong Kong 90% of SMEs that sought external help choose professional consultants. [17] While in the global study 63% of the companies cost in employing consultants was covered solely by enterprises. [14]

It is clear that two of the top four barriers ‘cost’ and ‘lack of environmental knowledge’ spring from the same problem that being lack of environmental management knowledge. Therefore addressing knowledge without the use of external consultants may provide a fundamental basis for an achievable EMS however further research into the methodology to achieve this is required. The lack of expertise in the area of environmental management is just one example of an area in which SMEs struggle. The characteristic well sighted for an SME, of having a lack of financial assets in relevance to their larger counterparts may originate from a lack of knowledge in many business areas thus having to seek costly external assistance. [3] Figure 3 lists other barriers that also feature within those studies listed.
ISO 14001 PROBLEMS & SOLUTIONS

Previous research has already highlighted the low uptake of ISO 14001 by SMEs and in particular by Irish SMEs. [18] [13] It is evident from this the standard has some defects. Although ISO maintain the standard is viable for all company types and sizes [3] the process of its feasibility, continual improvement, is a dubious process based upon an organisation’s own set objectives/targets. The process of ‘continual improvement’ has created much ambiguity amongst SMEs with McDonald 2003 noting the standard creates “…vagueness, as the work often relies on vague guiding principles of “continual improvement” [19]

ISO issued a guidance document to assist implementation, ISO14004. While the guidance was created for use by all size organisations ISO recognises, “the importance of SMEs and this guideline [ISO 14004] acknowledges and accommodates the needs of SMEs” [3] to some extent confessing the standards incongruity to SMEs. ISO 14001 was revised in 2004 and although some of the revised elements are easily understood, SMEs continue to have difficulty in;

1) auditing and identifying their environmental impacts
2) setting their objectives and targets, [3]

which are core areas in achieving continual improvement. These were the precise areas adapted from the successful ISO 9001 standard and blended into the revised 2004 version yet research reveals difficulties still remain. Other problems experienced is a heavy burden of paper work [14] [16] “a long paper trail exists in the certification process and many small companies do not have the time and staff resources in dealing with these onerous documents”. [20] This coupled with the lack of environmental awareness of staff [14] and sector specific tools depicts a losing battle for SMEs and provides ample reasoning in seeking third-party professional consultants. [15]

4. DISCUSSION

This section focuses on three areas, all of which are interrelated, where ISO 14001 overlook SMEs in achieving the standard. BS8555 a British standard EMS was developed to help businesses, SME’s in particular, to adopt an EMS in the most efficient and pragmatic manner, with an underlying objective to cause no upheaval to business activity. [6] The reasoning behind this EMS originated from the very low uptake and frequent failure of SMEs in implementing formal EMSs. [10] During its design inception it listened to the demand frequently expressed by SMEs for ISO 14001 to broken into various steps. BS 8555 drew from the success of ISO 14001 over EMAS noting that flexibility during implementation was of utmost importance for businesses. “EMAS is more rigorous than ISO 14001, it demands more”. [10] The workload of SMEs can fluctuate immensely in short time spans and they can be overburdened with work in the space of a couple of days. BS 8555 understood this and developed an EM system to adhere to fluctuating business workloads by incorporating phases of implementation. SMEs benefit from a phased system as it offers a withdrawal opportunity at anytime and in doing so provides external certification recognition of their efforts up to the phase of withdrawal. [21]

The strict onerous structured workload in ISO 14001 with no intermediate stages presents another problem. The success of many SMEs has been closely associated to the type of culture that exists within the particular SME. [22] Culture maybe described as what an average employee would feel inside the factory walls on a day to day basis. In the Organisational Science Journal, work culture was found firstly to be measurable and then to be directly related to important organisational outcomes, similar to that to the success of an EMS. [22] An SME in England stated that the BS 8555 allowed “A degree of flexibility which ensures the EMS works for the company – not the other way round” This flexibility generated by the stages in implementation allowed the particular EMS to become part of the ‘organisation’s culture’. [23] Pointing out that the 14001’s rigid work-intensive structure is detrimental in two ways. It firstly overloads the workloads of the employees attempting to implement the stand. While secondly its rigidity tightness an organisations culture reforming it to hamper and ultimately reject the management system. Unveiling the close association of ‘positive culture’ and ‘EMS success’ an idea which Kirkland and Thompson (1998) reiterate by suggesting corporate culture is important in determining the success of an EMS [24]
5. CONCLUSION

This paper has identified a sector that contributes significantly to negative impacts on the environment. SMEs are often overlooked when it comes to environmental finger pointing for a number of reasons. SMEs firstly are embedded in their local community and due to these close ties it is not surprising a level of acceptance is granted. While secondly owing to the nature of their business and size it is suggestive that little environmental impact occurs. Employing an average workforce of under 40 employees and manufacturing a renewable material, the TFM industry seem to impose negligible stress on the environment. However studies reveal that collectively the particular sector was found to be one of the poorest performing.

EMSs have been around for a number of decades their evolution into standards has unveiled ISO 14001 to be the most implemented and recognisable standard internationally. The standard initially tailored for larger companies and is not ideal for smaller enterprises. Indisputably many benefits are realised through EMS implementation. Case study examples have revealed Irish TFM to experience both internal and external benefits, while companies without any type of EMS suggest ISO 14001 would apparently offer little or no benefit at all to their business. This particular example underlines the key barrier of SMEs in first adopting, and secondly implementing an EMS a lack of knowledge on behalf of the SME. Comparison studies reveal ‘lack of environmental knowledge’ and ‘lack of finance’ as being the main obstacles in successful EMS implementation. Further investigation exposes ‘lack of finance’ is experienced through SMEs paying for professional consultants due to their …. lack of environmental knowledge!! When it comes to EMSs SMEs have a distinct lack of knowledge on the ‘environment’ and ‘management systems’. Unfortunately for SMEs ISO 14001 exacerbates these troubles.

The standard’s specific problems are four fold;
1) Auditing and identifying environmental impacts
2) Setting objectives and targets
3) Onerous Paperwork
4) Adherence to rigid structure

To accomplish the first two obstacles, SMEs currently turn toward external assistance a costly avenue for cash strapped SMEs. Case studies reveal seeking such professional assistance has been the thorn in many attempts of implementation due to the financial implications it presents. For SMEs to receive such assistance and tutoring without the use of costly consultants would represent not only a pragmatic answer but an intelligent one. Such a development is further complicated by the heterogeneous nature of SMEs. Environmental impacts for one industry may not be the same for another and therefore as a consequence their objectives and targets will also differ. On-going research similar in nature to the author, in developing sector specific tools to assist in EMS implementation provides the solution.

BS 8555 is a welcome development for SMEs in the field of environmental management. Tailored for SMEs it identifies the immense amount of paperwork necessary for ISO 14001 as a problem for the already multi-jobbed employees. Instead it exploits a workbook strategy that includes template sheets, case studies and checklists. The BS 8555 standard has seen immense growth in Europe with ISO’s technical committee overseeing ISO 14001, TC207, taking note of this growth and it’s suitability to SMEs in a recent meeting entitled ‘Be-aware! Small scale EMSs eventually might grow up.’

It is evident ISO 14001’s rigid structure does not accommodate SMEs, whose workload can fluctuate in such short time spans. Once implementation of ISO 14001 is undertaken it is all or nothing for SMEs with the latter occurring much too frequently. BS 8555 provides a route to ISO 14001 with a number of stages/phases on the way, providing an SME the opportunity to gain recognition up to the point of withdrawal. Positive culture has been identified as an integral attribute to EMS success. This structure allows the EMS to work for the company and not the other way around thus allowing it to become part of the organizations culture. EMAS is currently being revised with the new standard ready in 2010. The new standard will include a more objective/achievement orientated scheme and be simplified implementation into phases. This new model reflects the thoughts and findings in this research and shines a light on the future direction of Environmental Management Systems.
REFERENCES

Using biofiltration to clean waste gases

A M Gerrard*, J Skalicky and J Paca

*University of Teesside,
Middlesbrough, TS1 3BA,
England;

Institute of Chemical Technology,
166 28 Prague,
Czech Republic.

ABSTRACT
The paper gives an overview on the ways to clean waste gases using a variety of technologies. Emphasis is placed on biological processes. Some guidance is presented on how to model the behaviour of a biofilter used to remove styrene and acetone from air. The experimental data showed the inhibitory effect of the inlet acetone level and a new model to represent this is outlined.

INTRODUCTION

The cleaning of waste gases can be achieved by a number of physical and chemical approaches, including incineration, adsorption, scrubbing and condensation. Incineration may involve the consumption of extra fuel and in certain processes, may be catalytically enhanced. Adsorption onto solid beds of substances like activated carbon is also a possibility but usually the captured waste gas has then to be separated from the bed (by heating or a change in pressure etc.) Contacting the gases with a absorbing liquid is a common approach eg using caustic soda or amine solutions, but, as in the case of adsorption, regeneration may also be required. The condensation option is attractive when the gaseous waste stream is a valuable substance and a cheap cooling supply is available. However, there are a number of ways of approaching this worthwhile task using microbial action.

A biofilter is an example of such a process, see figure 1. It typically comprises a cylindrical or rectangular vessel through which the dirty gas passes. Within it, there is a solid bed of perhaps, natural material such as soil, compost, heather or activated carbon. However, plastic or glass packings having a high surface area are also commonly utilised. A colony of micro-organisms grows on the (moist) surfaces of these solids. As the gas mixture passes through the bed, the volatile organic compounds dissolve in the water and are then digested by the bacteria and/or fungi. Using these devices removal efficiencies of 90+% are possible, provided that the bed is sized correctly and the micro-organisms are kept in a healthy state.

*corresponding author: Tel: +44 (0)1642 342499; Fax: +44 (0)1642 342401; E-mail: a.m.gerrard@tees.ac.uk
The process works by the microbiological digestion of the odour-giving compounds as the gases pass through the packed bed. See reference [1] for an excellent summary of the subject. It is usually assumed that the VOC first dissolves in the liquid film surrounding the solid packing and then diffuses through it. As it does so, the resident microbiological organisms digest the pollutant, forming water and carbon dioxide. There is little cell growth in the bed, indeed if there is, then blockages can be a problem! The process is modelled by considering diffusion with bio-chemical reaction in the liquid phase, coupled with convection in the gas phase. The bio-reaction kinetics can be quite complex but they are often assumed to be zero or first order, (when analytic solutions can be found, see theory section). When the device is running in an unsteady state condition, then adsorption has to be included in the model.

**PRACTICAL DESIGN RULES**

If you are considering biofiltration as a possible solution to an odour problem, then the following ideas may be useful:

Low concentrations of VOC (typically tens to hundreds of ppmv) and high gas volumes favour this technology.

The overall removal efficiencies are often of the order 60-95 percent, hence if practically total removal is required, then this may not be an appropriate process. However, for easily biodegraded substances the efficiency, especially under laboratory conditions can approach 100%, (2). Mildenberger and van Lith [3] describe a very large biofilter at a sewage water plant which achieved 80% removal over a long period.

When removing sulphur compounds such as hydrogen sulphide, methanediol, dimethyl sulphide and others, good results have also been obtained using *bio-trickling filters*. (These are biofilters where there is a an additional continuous liquid stream passing down the bed to keep it wet at all times.) A reason for this is the much better pH control because of the circulating aqueous medium (the pH decreases due to the sulphuric acid formation in bio-oxidation process). The ability to keep the pH below 3 achieves almost total removal, even though the odour threshold concentrations of the sulphur compounds are around 15 ppb. In addition, the pressure drop across these beds are lower than those in biofilters due to different packing materials used, such as plastic Pall rings being used.
Two parameters are often used to get a rough sizing of the biofiltration unit. The first is the elimination capacity, EC, defined as the mass of VOC or odour removed per time per volume of bed. Appendix B of reference [1] suggests a range of values around 10-300 g m$^{-3}$ h$^{-1}$.

The other parameter is the empty bed residence time, $\tau$, which equals the volume of the empty bed divided by the volumetric flow-rate. This usually has a value in the range from a fraction of a minute to a few minutes. Because of the complexity of the biological and physical processes going on in the bed, the design process usually requires previous experience together with laboratory or pilot scale investigations.

The pressure drop across these beds is quite low e.g. a few centimeters water gauge, but it can rise if the bed collapses with age, or there is excessive cell growth.

A major operating variable to control is the moisture content of the filter, if the bed dries out too much or is too wet, then this threatens the viability of the micro-organisms present. (The feed stream can be humidified by spraying in water.) As mentioned before, the material used for the (moist) packing can be a natural material such as soil, peat or heather which have a naturally occurring microbial population. Activated carbon, perlite and plastic packings are also used, often with better mass transfer and pressure drop characteristics. The microorganisms can also be specially selected and optimised for a given separation task and are then introduced into the biofilter.

Broadly speaking, these devices work best when the conditions are steady i.e. flow-rate and inlet concentration remain within reasonable bounds. They can recover from sudden peaks and troughs, but this can take time. A more serious disturbance occurs when the supply of feed is cut off completely for a period. (There is then a danger that the microorganisms might die, if the interruption is sufficiently long.) Some systems require the periodic addition of nutrient solutions to supply the microbes with essential nutrients which ensure that they have sufficient catabolic activity for pollutant degradation, but this is usually not needed with natural packing materials.

On full-scale units, maintenance is needed to keep the spraying system in good order. Sometimes, the bed needs to be emptied and repacked to re-establish the original (lower) pressure drop and to redistribute the microorganisms. (There is some evidence that, after a while, there is a greater population of bugs at the inlet end of the filter and hence some of the rest of the bed volume may be running at a lower removal efficiency.

Overall, there is growing evidence [4] that a carefully designed and operated biofilter is capable of removing the majority of odour from a wide range of air streams including those found in water and sewage plants.

**THEORY AND RESULTS**

We shall now turn to a particular application of this process, the removal of styrene and acetone from air. Laboratory data collected in Prague has shown that both of these chemicals can be effectively removed by biofiltration but that the bugs have a marked preference for digesting the acetone rather than the styrene. Indeed, higher inlet concentrations of the more easily broken down acetone actually slows down the rate at which the styrene is removed. Simple models can be suggested to represent the removal of each substance. But better predictions of the performance of the biofilter arise when we also study the interaction between the two compounds. We have been able to model this behaviour, so that a rational way of sizing these devices can be attempted.

Ottengraf [5] has shown for “piston flow” conditions with no interfacial resistance that three analytic models may be derived. By assuming either “zero order” kinetics, (that means the rate of biological reaction is independent of the concentration of the pollutant concentration), then there are two cases here depending on whether the reaction or the diffusion is the slower (or “controlling”) process, the results are given by equations (1) and (2). His third model assumes first order kinetics, where the rate of reaction is proportional to the waste gas concentration, see equation (3). These three models offer a convenient way to linearise the expected concentration profiles, $C_G$ down the length of the bed, $h$.

\[
\frac{C_G}{C_{Gin}} = 1 - \left( \frac{h \cdot K_0}{C_{Gin} \cdot V_u} \right)
\]

(1)
\[
\frac{C_G}{C_{Gin}} = 1 - \frac{h}{v_a} \cdot \sqrt{\frac{K_v \cdot D_{ef} \cdot a}{2 \cdot m \cdot C_{Gin} \cdot \delta}}
\]

(2)

\[
\ln \left( \frac{C_G}{C_{Gin}} \right) = -\frac{h \cdot K_v}{m \cdot v_a}
\]

(3)

(The other parameters in the equations are usually combined into a constant whose value is found experimentally.)

For example, the new, pure styrene data was analysed first by studying linear, square root and logarithmic graphs of the measured concentration plotted against bed height. (The styrene concentration was measured at five points down the bed.) For a bed consisting of activated carbon, the best fit (i.e. highest coefficient of determination, \( r^2 \) and the lowest coefficient of variation of the rate constant (defined by the standard deviation divided by mean,) was given by the first order kinetics model. See figure 2 for an example, where the log-linear graph confirms the appropriateness of the first order assumption, using equation (3).

(For the perlite bed, zero order kinetics with diffusional limitation was slightly preferred, as shown by a good fit to a square root of concentration versus bed position graph.)

Then, the data on the mixture of styrene-acetone-air was studied to measure the interaction effect.

The experiments were organised with a constant air flow-rate and a constant inlet styrene concentration. Then the acetone inlet concentration was gradually increased from a low to a high level. There is clear experimental evidence that the two VOC’s (styrene and acetone) influence the rate of removal of each other. The removal efficiency of (the more difficult to digest) styrene progressively declined as the acetone inlet concentration increased. The acetone removal also declined with its increasing inlet loading. See figure 3.
Following a similar approach to the pure styrene analysis, first order kinetics seemed to fit best for both the styrene and the acetone removal rates when each was modelled independently of the other. However, there was a steady decrease in the value of the fitted rate constant (for both styrene and acetone) as the inlet acetone concentration increased. This interesting behaviour can be modelled as follows.

**NEW MODEL**

If we assume the acetone follows first order kinetics with some self-inhibition then:

$$\frac{dCa}{dh} = -B_1 \cdot Ca \cdot (1 - p \cdot Ca)$$  \hspace{1cm} (4)

where $p$ is the inhibition constant.

This can be solved to give:

$$Ca = \frac{Ca_{in} \cdot \exp(-B_1 \cdot h)}{[1 - p \cdot Ca_{in} \cdot (1 - \exp(-B_1 \cdot h))]}$$  \hspace{1cm} (5)

The digestion of the styrene is also assumed to be first order and to be inhibited by the presence of the more easily digested acetone (which is preferentially consumed by the micro-organisms.)

$$\frac{dCs}{dh} = -B_2 \cdot Cs \cdot (1 - p' \cdot Ca)$$  \hspace{1cm} (6)

This can be integrated:

$$Cs = Cs_{in} \cdot \exp\left(-B_2 \cdot \int \left[1 - \frac{p' \cdot Ca_{in} \cdot \exp(-B_1 \cdot h)}{1 - p \cdot Ca_{in} \cdot (1 - \exp(-B_1 \cdot h))}\right] \cdot dh\right)$$  \hspace{1cm} (7)
(The remaining integral in equation (7) can be solved analytically but it is more convenient to solve it numerically using the Trapezium rule.)

Then the parameters $B_1$, $B_2$, $p$ and $p'$ were altered to give the best fit to the measured concentration profiles of acetone and styrene down the bed’s length. The following table gives the best values found.

### Table 1: Optimal values of $B_1$, $B_2$, $p$ and $p'$ parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_1$</td>
<td>3.6</td>
</tr>
<tr>
<td>$p$</td>
<td>0.00083</td>
</tr>
<tr>
<td>$B_2$</td>
<td>3.85</td>
</tr>
<tr>
<td>$p'$</td>
<td>0.00093</td>
</tr>
</tbody>
</table>

When these values are put into equation (7), then the performance of the biofilter can be modelled. Figure 4 shows the actual and predicted data.

![Figure 4](image_url)

Figure 4 is a scatter diagram to show the adequate fit of this model to the experimental data.
CONCLUSIONS

Biological processes are well worth considering for cleaning industrial waste gases. Usually practical experience and laboratory data are needed to design these devices. The new experimental data on the styrene-acetone-air system showed that the acetone inlet concentration affected both its own removal rate and that of styrene. A new model has been proposed to describe this process.

NOMENCLATURE

- $B_1, B_2$: kinetic parameters
- $C$: gas concentration, g.m$^{-3}$
- $C_{i}$: inlet gas concentration, g.m$^{-3}$
- $h$: filter bed height, m
- $p, p'$: inhibition parameters

subscript
- $i_{in}$: inlet condition

ACKNOWLEDGEMENT

This study was financially supported by EU Socrates scheme which funded Mr Skalicky’s visit to the University of Teesside. Experimental work was supported by the Ministry of Education of the Czech Republic, Project MSM 6046137305.

REFERENCES


Voluntary Environmental Instruments - their benefits and usage in the Czech Republic

Tereza Kadlecová (nee Fidlerová)¹*, Lilia Dvořáková², and Marcela Srchová³

¹Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

²Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

³Department of Industrial Engineering and Management
University of West Bohemia
Pilsen, 396 14, Czech Republic

ABSTRACT

Supply enormously outstrips demand in the current corporate environment, market competition is growing and the struggle for customers is growing in intensity. Companies have to face up to new and even stricter environmental legislative requirements. Steadily increasing prices of raw materials and energy are another threat for many companies. Last but not least higher customer requirements have to be satisfied. Companies need to differentiate themselves from their competition. However, one can differ only in what is not obligatory for others, i.e. what is voluntary. Some companies have already learnt that there are some activities not required by regulations that, despite being voluntary, are not useless. Voluntary environmental activities can become a significant competitive advantage factor which creates market differentiation.

1. INTRODUCTION

1.1 PREVENTIVE VERSUS REACTIVE STRATEGIES

Many forms of preventive as well as reactive strategies for environmental conservation are being used in industrial engineering.

Cleaner production, ecodesign and ecolabelling are examples of preventive strategies. As the name indicates the preventive strategies endeavour to prevent the origin of damage, they seek for sources of pollution and waste. Preventive strategies have more potential and their realization should be supported. The reason why reactive strategies are not so effective or promising is the fact that they do not focus on the sources of environmental damage, they only try to mitigate the consequences of production. Among these “end-of-pipe” technologies are, for example, refuse compactors, collection containers and vehicles, waste heat recovery systems, air pollution filters, noise abatement investments and sewage treatment plants. As a result the quantity of toxic agents in one environment domain drops, but rises in another domain. As practice has shown, the preventive strategy has a positive impact on the environment and it leads to financial savings, economic profit, cost reduction and enhancement of the competitive advantage at the same time. Preventive strategy is therefore considered as a double profit strategy: environmental and economic – a so called ‘win-win’ strategy.

Companies have a lot of voluntary environmental instruments at disposal. There are two types of voluntary environmental instruments: regulative and informative. Among the most obvious voluntary instruments are cleaner production, ecodesign and ecolabelling.

2. CLEANER PRODUCTION

2.1 CONCEPT OF CLEANER PRODUCTION

¹* Tereza Kadlecová: Tel.: +420 377 638 464; Fax: +420 377 638 402; E-mail: fidlerot@kpv.zcu.cz
The concept of cleaner production is being connected with the integral preventive strategy which is applied especially to the sector of production. The aim of this strategy is to do away with the causes of environmental pollution as a consequence of production.

Within cleaner production the material and energy flows are monitored in order to identify the sources of undesirable waste. Afterwards the possibilities of elimination of these sources are explored, namely:

1. Ease of technical feasibility;
2. Final economic efficiency;
3. Environmental efficiency.

All the company processes are observed as a whole - in terms of impact on all domains of the environment. It is therefore not possible to transfer the negative impact from one domain of the environment into another one, as it is with the end-of-pipe technologies. The implementation of cleaner technology in a company is not a one-shot action, but it is a long-term process. Effectiveness is the reason why a company should deal with cleaner production. In other words, reducing raw material and energy consumption reduces the negative impact on the environment. Waste increase is thus being prevented at the source and this leads to a significant economic effect at the same time.

The main features of cleaner production are:

- preventive approach to solving the problems;
- integration principle;
- wide range of application (processes, products, services);
- economic benefits.

Type of product, technology used, machines and equipment, raw material, work organization and adherence to the manufacturing processes are all controlled as part of cleaner production.

2.2 CLEANER PRODUCTION SPHERE OF INTEREST

Cleaner production, as a preventive strategy for environmental conservation, focuses on searching for the origin of waste. In the cleaner production main sphere of interest are:

- product character,
- technology used,
- machines and equipment,
- raw materials,
- adherence to manufacturing process principles,
- work organization,
- employees approach to the tasks,
- company management system.[1]

In the above examples the impact on the environment can be reduced by engineering as well as non-technical (organizational) methods, which are very effective and are no or very low costs in many cases.

One can prevent waste origin by:

a) better production logistics and work organization – this is the most convenient economic option (e.g. careful treatment with the raw material,

b) technology changes – simple machines adjustments, changes in methods principles,

c) change in main material,

d) product chase – e.g- reduction of materials used for production,

e) waste restoration in the same process where it arose, so called reverse waste,

f) waste recycling in the same company.[1]
2.3. CLEANER PRODUCTION IN THE CZECH REPUBLIC

Cleaner production activities began in the Czech Republic in 1992. However, the application level of cleaner production is in practice rather low, as can be seen in the following figure.

![Figure 1: Number of cleaner production projects in years 2002 - 2004](image1)

3. ECODESIGN

3.1. CONCEPT OF ECODESIGN

Ecodesign is one of the preventive oriented voluntary regulative instruments of environmental policy. While cleaner production focuses on a company as a whole, the concept of ecodesign concentrates on product development and design.

Ecodesign incorporates requirements of environmental protection into product design and development. So far there is no unified definition of ecodesign. In general, ecodesign can be defined as a systematic process of product design and development which puts emphasis not only on classical features such as functionality, economics, safety, ergonomics, technical feasibility, aesthetics, but it also puts emphasis on the minimum negative impact of a product on the environment.

![Figure 2: Influences on product design and development, [4]](image2)
The main purpose of ecodesign is to design a product with a minimum impact on the environment in its own life cycle (“from cradle to grave”).

Within ecodesign all phases of the product life are considered, especially:
- getting raw material;
- production;
- usage;
- disposal.

This means that the method of disposal is considered during product design. It is therefore the most preventive and economical approach towards the mitigation of a product’s negative impact on the environment.

Outlines of ecodesign are as follows:
- enforcement of safe products and services,
- preservation of biosphere,
- sustainable usage of natural sources,
- minimize waste and increase recycling,
- wise energy utilization,
- reducing environmental and health hazards to employees and customers,
- exchanging of information among the eco-designers. [5]

### 3.2 TYPES OF ECODESIGN STRATEGIES

1. Development of a brand new product concept
2. Change in product construction,
   a) Strategies focusing on product ingredients,
   b) Strategies focusing on product structure
   c) Strategies of waste disposal optimization [5]

Ecodesign is not an unknown concept for Czech engineering designers and some manufacturing corporations. Ecodesign activities are mainly concerned with:
- substitution of dangerous material with less dangerous ones,
- implementing measures leading to waste minimization,
- reduction in raw material consumption,
- packing and transportation optimization.

### 3.3 ECODESIGN IN THE CZECH REPUBLIC

Product innovation volume in the Czech Republic, however, is unsatisfactory. On the basis of the Innovation and Development Centre investigation from 2004 it is obvious that the total investment in development and innovative technologies in companies was only 48 billion Czech crowns, which is less than 2 % of the sales in all innovating companies.

According to these investments 45 % was spent on new technologies and equipment, but only 2 % was invested in design projects.

### 4. ECO-LABELLING (ENVIRONMENTAL DECLARATION TYPE I.)

#### 4.1. CONCEPT OF ECO-LABELLING

Ecolabelling is a certification system for products and services that are friendlier to the environment than similar products and do not have such a negative impact on the environment. This system is directed by a third party who has to be independent.
Nowadays there are more than 30 ecolabelling systems and their number is increasing. National or supranational labels are assigned within those systems. Ecolabelling is regulated by ISO 14024.

The main task of ecolabelling is to:
- choose product groups whose negative impacts on the environment can be reduced,
- set a group of requirements that are to be met by the product groups, so that the product groups can be legitimately considered as environmentally friendlier. These requirements are to be disclosed in an appropriate way.
- assign the eco-labels (in the form of a licence contract) to products that meet the given requirements,
- crosscheck the compliance of features of the labelled products with the licence contract requirements. [2]

The products and services that apply for certification have to meet many requirements concerning the quality of the product (service), particular production phases, and usage of raw material, technologies and final disposal.

In contrast to the majority of preventive strategies focusing on systematic examination of manufacturing processes, ecolabelling makes use of market mechanisms outside the company. The basis of market mechanisms is supply and demand. Demand is created by consumers interested in environmental issues and choosing environmentally friendly products for their consumption.

Not only features of a finished product, but all stages of product origin are assessed, despite the fact that ecolabelling is an instrument related to a product. By means of ecolabelling it is possible to influence the usage of raw material, choice of technologies as well as disposal methods of used products.

Ecolabelling is the oldest voluntary instrument; it originated in the Seventies in the then Germany. The first eco-label was ‘Blue Angel’ and it was awarded in 1978. Nowadays ‘The Flower’ is the trademark of ecolabelling in the EU.

![Trademark of ecolabelling in the EU – ‘The Flower’](image)

### 4.2. ECOLABELLING IN THE CZECH REPUBLIC

In the Czech Republic the ecolabelling system is executed as the National Programme of Environmentally Friendly Products Labelling, respectively on the basis of the Government decree Nr. 159 from April 7, 1993. [3] The following figure shows the number of ecolabelling licences awarded in years 1994 to 2004.
Despite this, the environmentally friendly products market has so far not evolved very much in the Czech Republic. Economic return is therefore relatively low. Even though the fees the producers pay are much lower than for ecolabelling systems in other countries, the ecolabelling eco-efficiency is rather poor.

5. BENEFITS ARISING FROM ENVIRONMENTAL FRIENDLY BEHAVIOUR

Many companies are reluctant to implement environmental management because they are afraid of enhanced costs connected with environmental preservation.

One could say that the application of voluntary preventive strategies and instruments is not worth it. It is very costly to monitor energy and material flows, environmental costs and revenues. Using environmentally friendly raw material and other resources brings with it higher costs. Are there any benefits resulting from a company’s environmental approach?

Measures for environmental protection can really increase costs in the short-term. However in the long-term those measures should bring a variety of financial as well as non-financial benefits to a company. The question is what are the company’s preferences - short-term profit or fixing the market position, building reputation and long-term sustainable growth?

There are two ways of looking at the benefits arising from a company’s environmental behaviour and application of voluntary instruments. The first is from the benefit point of view— benefits can be divided into financial and non-financial benefits. Benefit origin is the other point of view. Benefits can be internal (arising in a company) or external (flowing from outside the company). The table below displays benefits arising from environmental behaviour divided according to these four criteria.
<table>
<thead>
<tr>
<th>ORIGIN POINT OF VIEW</th>
<th>EXTERNAL</th>
<th>INTERNAL</th>
</tr>
</thead>
</table>
|                      | • Lower fees and fines for polluting the environment  
|                      | • Larger market share acquisition  
|                      | • Lower production costs connected with waste reduction and higher efficiency (e.g. material and production costs savings)  
|                      | • Lower environmental damage remedy costs  
|                      | • Improved trading income as a result of lower costs  
|                      | • Waste disposal costs savings (incl. reduction of investment and expenditure for running end-of-pipe technologies)  
| FINANCIAL            |          |          |
|                      | • Decrease in environmental risk  
|                      | • Improved position of a company during negotiation with public institutions  
|                      | • Enhancement of company image  
|                      | • Simplification of EMS  
|                      | • Getting into the new markets  
|                      | • Possibility of engaging in a net of other environmentally conscious companies  
| NONFINANCIAL         |          |          |
|                      | • Lower waste production (fulfilment of legislation requirements concerning preservation and improvement of the environment)  
|                      | • Simplified waste management due to lower requirements imposed on storage areas for waste, waste transport and equipment for waste disposal  
|                      | • Improvement of working environment  
|                      | • Possible improvement of product quality  
|                      | • Including and interesting employees, possibility of getting first-rate employees  
|                      | • New technological innovation  

6 SUMMARY

Nowadays more and more emphasis is put on the protection of the environment. Companies can be voluntarily environmentally active above the ambit of the legislation. Companies have a wide range of voluntary environmental instruments at their disposal. One could say that applying voluntary preventive strategies and instruments is not worth it. There are a lot of financial as well as non-financial benefits arising from companies’ environmentally friendly behaviour. If a company wishes to be ahead of the competition it has to do more than the competition, and do what is not obligatory. However, the level of usage of these voluntary instruments is, for various reasons, unsatisfactory in the Czech Republic. Among the main barriers are poor awareness of these instruments, aversion to the new, the insufficiently developed green-products market and finally, lack of funds.
REFERENCES


[2] Ecolabelling, Ministry of the Environment in the Czech Republic

[3] Aid programme for environmental technologies in the Czech Republic Program, Ministry of the Environment in the Czech Republic, Prague, 2005


Manufacturing of a Pick-and-Place Robot with Inverse Dynamic and Robust Controllers

S. Saeed Mirian¹, S. Mohsen Safavi²

¹Islamic Azad University
Najafabad branch, Isfahan, Iran

²Isfahan University of Technology
Isfahan, Iran

ABSTRACT

In this research, designing and manufacturing of a pick-and-place robot, along with its kinematic and dynamic analysis have been introduced and surveyed. Simulations of the robot’s movement and control procedure via nonlinear robust and inverse dynamic controllers are then discussed as well as the comparison of uncertainties among them as modern control methods. The main function of this robot is coating epoxy powder on “Hawle” gate valves at temperatures up to 200° C. The results obtained in this study can greatly fulfill the need to resolve industrial problems.

1. INTRODUCTION

Pick-and-place robots are the most widely used traditional robots in industry with their functionality covering a broad range of tasks. The primitive type of these robots was first designed and assembled in 1967 for a precise casting process. The main reason to employ such mechanism at that time was hazardous environment for human operation considering toxic, gaseous emissions from the melt along and also the required high speed operational cycles and complex handling, which makes up the most prominent incentives of attention to industrial robots in the frame work of human factors engineering [1]. Since the first types of these robots were fabricated, world wide researches have been carried out on design, construction, handling control, path optimization, and their dynamic load capacity enhancement. Among such researchers, Fateh [2] has examined Fuzzy Impedance control in rapid operation of robots based on position, velocity, and force feedback of end-effect with the purpose of overcoming uncertainty induced by external disturbances, friction, inertia, and other dynamic effects. Moosavian and Homaeinejad [3] have studied applying of sliding mode control on a 2-DOF manipulator and reduction in chattering frequency that causes reduction in energy consumption. Ahmadi and Eghtesad [4] applied Image processing approaches to design and control of a catcher robot. Qurayem and Nickoubin have presented an algorithm to calculate dynamic load carrying capacity of a manipulator with optimal control through predetermined path. Wu et al [6] have studied an robust controller based on SISO nonlinear systems in order to precisely control the displacement of linear motors of a two dimensional table. Moreover Chon and Kwan [7] have investigated modern post processing methodologies for off-line robot programming within CIM system and reached to acceptable conclusions. This robot is a 3-DOF mechanism, which its primary function is loading Hawel valves of any type and size up to a specific capacity, handling, coating and finally unloading them on an index table at the end of each operation cycle. Securing the required level of stability in any kind of circumstances by using robust controller, considering the variable loading capability, minimizing energy dissipation, enabling off-line changes into the program and movement path, facilitating modifications in system operation and accuracy of the gripper trajectory, constitute the outstanding advantages of this robot. The results of this study are widely used for handling mechanisms, automation and computer aided process planning. Following sections involve the robot performance in a motion cycle based on its mechanical structure, simulating and designing of robot motion in one optimal path.

* Corresponding author: Tel.: (+98) 311-3915238; Fax: (+98) 311-3912628; E-mail: saeedmir@yahoo.com
cinematic and dynamic analysis, applying of inverse dynamic and robust control systems and introducing feedback parameters. The induced errors in each cycle are compared and discussed in the results.

2 DESIGNING AND MODELING OF ROBOT

The robot with a structure depicted in figure 1 which consists of about one-hundred mechanical, electronic and pneumatic parts, is able to start moving from a base position and coat epoxy powder on the Hawle valves with an angle of about 120° related to their initial position (θ=0°) after gripping them from an index table or conveyor belt in a cycle regarding to possible linear movements of the system (P), the hinge movement of the arm (R1) and the rotational movement of the gripper (R2), in favorable conditions. This action takes place according to figure 1 and figure 2 until the linear moving body reaches the first defined point within the length limitation of linear movement (L−0) and after the reduction of the arm angle with the horizon until it reaches to its ultimate limit in downward direction (α=36°). The angle at which the valve enters the coating tank is the angle with the best condition for absorbing the paint by the valve with the least loss (θopt=120°). After the valve is immersed in the coating tank for a definite time, the valve leaves the paint by a command, simultaneous with the re-ascension of the arm and repeats the movements as in the previous stage but in reverse direction to return the coated valve to its initial position (L~0.5, θ=0°). While the Hawle valve is leaving the coating tank, the gripper is required to accompany the rotational movement with vibration by a stepper motor in order to prevent the loss of high expenses coating materials as much as possible. While an index table or conveyer belt is employed as part feeding device, the robot is ready to perform operations on the second valve after the transfer mechanism is moved for one step.

Figure 1: Real and schematic views of the robot

Figure 2: Degrees of freedom: P (Linear Motion, L), R₁ (Arm Rotation, α) and R₂ (Gripper Rotation, θ)
Weight control of the parts and compensation of the pneumatic actuator weakness in achieving the design of every robot, the design algorithms of the movement path and robot function are the most important stages on the basis of which the required parts are designed or selected [4]. After considering different factors such as the best controlling methods for the parts of the robot to reach a complete and uniform coating, the minimization of power loss and maximum use of existing space by deleting the redundancies of degrees of freedom in such robot, studying and performing the required calculations on its initial design depicted in figure 3(a). The ultimate structure of robot was finally modified to a satisfactory final design configured in figure 3(b). The most important parameters employed in the design and simulation of robot are weight and dimensions of the Hawle valves and robot consumption power as variable quantities on the one hand, and end-effect position at intervals between movement and stopping time in one cycle as the specified quantities on the other hand. All performed through the controllers under study in this research by angle/torque feedback system, lead to successive movements of the intended parts in the specified path with desired precision in uncertainty conditions whether by sudden changes of the load or by pneumatic system weakness. The consumed power of the system which is drastically the function of time of each cycle will reach to its least amount by optimization of the movement path and quantification of time.

3 Kinematic and Dynamic Analysis

In order to facilitate the analysis of this system, the robot degrees of freedom were separated into one linear movement and two rotational movements related to one robot with two degrees of freedom and the results obtained were combined with each other. Therefore, in this section only the process of cinematic relations of one robot with two degrees of rotational freedom is discussed. In the cinematic analysis of this problem, the rotational matrices of A-1 is specified in the form of Equation 1 and the solution for finding its rotational matrices is performed by employing Denavit-Hartenberg amounts in table 1 [8, 9, 10].

In this table different quantities such as $\theta_i$ in rotational manner from $X_{i-1}$ to $X_i$ around $Z_{i-1}$, $\alpha_i$ rotation from $Z_{i-1}$ to $Z_i$ around $d_i$, $X_{i-1}$ distance from $X_{i-1}$ to $X_i$ in the direction of $a_i$ and $Z_{i-1}$ the distance from $Z_{i-1}$ to $Z_i$ in the direction of $X_{i-1}$ are defined according to figure 4. In this figure, the two-link system of robot is specified by specifying the main link and its mass center position on relative coordinates related to the center of rotation and in figure 5 the physical specifications related to the main link such as material, mass, and moment of inertia rates considered in dynamic calculations are taken into consideration [8, 9, 10].
The unknown angles of $\theta_i$ can be calculated by knowing the final position of the gripper and by employing Eqs 2 and 3 inversely. Also Eqs 4 and 5 are used in this article in order to solve the gripper instant velocity and acceleration by Jacobian Matrix (JM); so that the numerical data extracted from the following equations will be employed for the processing related to the path follow-up in the control simulation and will be pointed in next sections [10].

**Table 1**: Denavit – Hartenberg parameters

<table>
<thead>
<tr>
<th>Link</th>
<th>$\theta_i$ (rad)</th>
<th>$\alpha_i$ (rad)</th>
<th>$d_i$ (m)</th>
<th>$a_i$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link 1</td>
<td>$\theta_1$</td>
<td>$\pi/2$</td>
<td>0</td>
<td>$l_1 = 794.7$</td>
</tr>
<tr>
<td>Link 2</td>
<td>$\theta_2$</td>
<td>0</td>
<td>0</td>
<td>$l_2 = 283.1$</td>
</tr>
</tbody>
</table>

\[
A^0_1 A^1_2 A^2_3 A^3_4 A^4_5 \ G = X
\]  
\[T^0_2 = A^0_1 A^1_2 = \begin{bmatrix}
\cos \theta_1 \cos \theta_2 & -\cos \theta_1 \sin \theta_2 \\
\cos \theta_2 \sin \theta_1 & -\sin \theta_1 \sin \theta_2 \\
\sin \theta_2 & \cos \theta_2 \\
0 & 0
\end{bmatrix}
\]  
\[
\sin \theta_1 \ - \cos \theta_1 \ l_2 \cos \theta_1 \cos \theta_2 + l_1 \cos \theta_1 \\
- \cos \theta_1 \ l_2 \sin \theta_1 \cos \theta_2 + l_1 \sin \theta_1 \\
0 \ l_2 \sin \theta_2
\]  
\[\begin{bmatrix}
\sin \theta_1 \\
- \cos \theta_1 \\
0 \\
0
\end{bmatrix} \]  
\[\begin{bmatrix}
\cos \theta_2 \\
\sin \theta_2 \\
0 \\
1
\end{bmatrix}
\]  
(1)  
(2)
With regard to Eq. 6 in which $\tau$ is the moment vector, $\mathbf{M}(\mathbf{q})$ is the inertia matrix, $\mathbf{C}(\mathbf{q},\dot{\mathbf{q}}) \in \mathbb{R}^{n \times n}$ is the matrix of centrifugal force effect and $\mathbf{g}(\mathbf{q}) \in \mathbb{R}^n$ is the effect of gravity in the dynamic analysis of the robot, then it is possible to specify the torque related to drive motors. The rates of torque obtained here along with the obtained empirical rates from the feedback system are analyzed and compared which will be pointed out in next sections [8, 9, 10].

$$\tau = \mathbf{M}(\mathbf{q}) \ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) \dot{\mathbf{q}} + \mathbf{g}(\mathbf{q})$$

$$\{\ddot{\mathbf{x}}\} = \left[\mathbf{J}\mathbf{M}\right] \{\ddot{\mathbf{q}}\} + \left[\mathbf{J}\mathbf{M}\right] \{\dot{\mathbf{q}}\}$$

4 Applying Control on the System

4-1 Inverse Dynamic Control

In the control of robots, the Inverse dynamic control is one of the control methods employed in non-linear systems with optimized non-sympathetic stability. In this section vector $\mathbf{g}$, moment of inertia, $\mathbf{m}$ and $\mathbf{z}$ are first calculated and stored. With regard to the Inverse dynamic control law which is defined in equations 7 and 8 for the robot and equation 9 for the controller, $U_0$ is replaced in the first line of equation 9 and equation 10 is resulted, which can be compared with equation 7 and the general law for this control is extracted according to equation 11. [8, 9, 11]

$$\tau = \mathbf{M} \ddot{\mathbf{q}} + \mathbf{C} \dot{\mathbf{q}} + \mathbf{g}$$

$$\tau = \mathbf{M} U_0 + \mathbf{C} \dot{\mathbf{q}} + \mathbf{g}$$
With regard to the difference in the function of $\theta$, the rate changes related to each $\theta$ is inevitable, therefore, $K_p$ and $K_d$ matrices in this research with regard to the robot specifications and constants, are considered the diametric matrices with different rates on diameters. The application of such controller requires inertia matrix, Coriolis and centrifugal vectors, gravity acceleration and system damping calculations. These quantities required to be calculated in on-line manner, because in this position, the control is based on non-linear response in the present condition of the system, therefore, the mentioned calculations of the quantities are not possible before performing operations and in off-line manner. The Inverse Dynamic control laws also require that the system dynamic model parameters be recognized precisely and movement complete equations be calculated in real time. The intended model is usually recognized on the bases of incomplete knowledge present in relation with mechanical parameters and un-modeled dynamics along with a degree of indefiniteness. Also, the unknown rate of dependence of the model to the rate of end-effector load and its load carrying capacity causes not be able to define an appropriate compensation rate in the controller for it. As a result, the use of other controllers which lack such a limitation are taken into consideration. Even though this control method is able to provide an appropriate function, but compared to the moment disturbances and environmental indefinites, it enjoys high sensitivity. As a result, a robust controller can be employed for overcoming such problems. This controller, in addition to providing logical responses, enjoys reliability, strength, and appropriate stability in the presence of moment disturbances and uncertainties. [8, 9, 11]

4-2 ROBUST CONTROL

The aim in robust control is to estimate few parameters in online manner instead of estimation of large number of independent dynamic parameters of a robot. In this position the $M, \dot{C}, \dot{\bar{g}}$ terms are not available and deleted from control command. By taking equations of robust controllers, a law will be extracted according to equation 12 [12, 13]. After simplification, Eq. 13 will result in which “$w$” is replaced according to Eq. 14 and robust control equation will be obtained in the form of Eq. 15 [10, 12, 13].

\[
\tau = -K_d \dot{\sigma} + u_0 = K_d r + u_0
\]

\[
u_o = -\left(\beta^T \alpha^*\right)^2 \sigma
\]

\[
u_o = -\left(\beta^T \alpha^*\right)^2 \frac{\beta^T \alpha}{\beta^T \sigma} \sigma
\]

\[
\dot{\sigma} = -K_d \sigma
\]

\[
M \dot{\sigma} = -K_d \sigma + u_o - C \sigma - w
\]

\[
w = M (q) (\dot{q} - \dot{\bar{q}}) + V_m (\ddot{q} + \dot{\bar{q}}) + g (q)
+ F_d \ddot{q} + F_s (\dot{\bar{q}}) + \tau_d
\]

\[
M \dot{\sigma} + C \sigma + K_d \sigma = u_o - w
\]

\[
\rho \geq \|w\|
\]
In this article $\rho$ is calculated as a scalar number with the estimation of $\delta$ online according to equation 16.

$$\rho = \delta_0 + \delta_1 \|e\| + \delta_2 \|e\|^2 \geq \|w\|$$  \hspace{1cm} (16)

Figure 6: Tracking the path with considering uncertainties via

(a) Inverse Dynamic control and (b) Robust control.

5 Analyzing the results of the trend of two control methods

After simulation of the two Inverse Dynamic and Robust control methods by MATLAB software on the basis of the data obtained from the dynamic and kinematic analysis of the robot, the results of this section were studied and compared. It is assumed here that the real $q_r$ are different from $q_d$ by 0.005 radians and the speed and acceleration are different from the given data by 0.001. In cases where there is no uncertainty, the resultant results of the Robust controller with Inverse Dynamic controller are completely similar and the error is quickly biased to zero in following a desired path. In uncertainty conditions in the system, the results obtained from comparison of Robust with Inverse Dynamic are different as depicted in figure 6. As the chattering phenomenon occurs in the Inverse Dynamic method, the error value compared with robust controller is higher. This phenomenon is specified in figure 7. It shows how the problems in obtaining the required precision in tracking an optimized path against obstacles are removed by utilizing robust controllers. [12, 13] The most important cases considered here as the uncertainties in the controller, are the probable changes in weight and dimension of the parts, sudden loads, and weakness in the pneumatic actuator system.

Figure 7: The error value and the manner of chattering phenomenon
6 Feedback Systems

There are different systems and quantities which are employed in many experiments as the bases for hardware and controller feedback variables. Attention to the angle feedback by exploiting rotary encoders is a common method for this job in which the extracted data from this method in following the robot movement path within the time of one cycle is specified in figure 8. Although the basis of angle feedback is more precise comparing to the other methods such as torque feedback, the obtained instant responses which are considered as the controller inputs for evaluation of the rate of uncertainties are consistent to each other in appropriate conditions within the limited defined tolerances and only within short time limits. In conditions that load changes produce sudden disturbances resulting in great torque changes in the system, noticeable differences are observed. In this state, due to chattering phenomenon, Inverse Dynamic controller is not able to keep the required precision. On the other hand, the use of the torque method provides the possibility of proper adjusting of the frequency so that it is possible to adjust the times of activating acceleration and stoppage of the system in a way that the uncertainties resulting from inertia reaches to its least amount and its bias to zero is neglected.

7 Conclusions

In this article, after pointing to mechanical structure and function of a pick-and-place robot, simulation of robot movement in an optimized path, kinematic and dynamic analysis, we embarked on the trend of the two methods of control as Inverse Dynamic and Robust methods in following a desired path due to simulation and experiment analyses and also results obtained from the error values in each cycle were compared by means of introduced feedback systems and quantities. Consequently the use of a robust controller seems necessary since the change in mass or size of the parts displaced by the robot and other probable uncertainties should be considered.

References

Manufacturing of a Pick-and-Place Robot with Inverse Dynamic and Robust Controllers

Pages 1375-1383, 2005.


A Survey on Impacts of Shift Work to Shop Floor Workers’ Safety and Health in a Manufacturing Company

Baba Md Deros*, Nor Kamaliana Khamis, Ahmad Rasdan Ismail, and Asri Ludin

Department of Mechanical and Materials Engineering
Faculty of Engineering & Built Environment
Universiti Kebangsaan Malaysia
Bangi, Selangor, 43650, Malaysia

ABSTRACT

This survey discusses the impacts of shift work towards shop floor workers’ safety and health in a manufacturing company. This study had used the survey methodology to study the shop floor workers’ perception of shift work towards their safety and health. The survey questionnaires were distributed to 200 shift workers in the production department. The survey questionnaire covered in detail the shift workers’ background, health and safety effects and personal factors towards safety and health. The survey was conducted in a company, which is practising a rotating 3-shift, 8-hour per shift system. In order to identify the most dominant factor that influence safety and health among shop floor workers, the survey results were analyzed using the descriptive and correlation technique. Based on the findings obtained from the survey, the authors had suggested the case study company to be more sensitive on the effects of shift working towards shop-floor workers’ safety and health.

1. INTRODUCTION

The term ergonomics is derived from the Greek words ergon (work) and nomos (natural laws). Ergonomics is the study of human capabilities, physical and mental limitations in relation to their work demands and fulfilling the natural laws [1]. Shift-work is broadly defined as an arrangement of working hours that uses two or more to extend the operational work hours beyond the conventional office hours [2]. Shift-workers might work in the evening, in the middle of the night, overtime or extra-long workdays. They might also work regular days at one time or another. These types of schedules (i.e. combination or rotating shifts) are commonly practice in manufacturing industry [3].

2. LITERATURE REVIEW

In today’s modern technology, it had made possible to perform many activities at any time of the day or night. Our “24-hour society” requires that important services be provided at all times [3]. Shift work is widely used in manufacturing industry for economical reasons, such as increasing capacity rate to more than 3 times, compared to the normal 8-hour shift production output. This system is required by technological requirements for continuous plant operation that need to fully utilize the available machines and tools [4].

Bernardino (1713) as quoted by Costa (1996) states that, although the benefits of rotating shift to increase productivity and decrease production cost, but on the other hand, it can present high risk to human [5]. In fact, it can have negative impacts on workers’ health and well-being, in four aspects such as: biological function disruptions, working ability, social interaction, and deterioration of health. Biological function disruptions are caused by disturbance of normal circadian rhythms of the psycho physiological functions, beginning with sleep and wake cycle. Logically, we expect the impact of multiple shift work features shall reduces workers’ productivity and results in higher accident and a mistake rate during work. Difficulties in social interactions such as maintaining the normal relationships both at family and social level, which result in negative influences on marital relationships, care of children and social contacts [6]. Barnes-Farrell et al. [7] believed that erratic shift can cause conflict of interest between work and family.

* Corresponding author: Tel.: (603)-8921 6117; Fax: (603)-8925 9659; E-mail: hjabba@eng.ukm.my
At the same time, they can also contribute to other potential risks such as: psychology condition, digestion disorder, short sleep length, and physical system such as anxiety and interference. Furthermore, in the long term, it could cause problems with respect to gastrointestinal, neurophysic, cardiovascular control and sadness [6].

3. Research Methodology

Data collection involved two main sources namely: a survey method and a case study on the design of shift work schedule. The data collection techniques used was observation method, structured interviews and analysis of document. Data analysis had used descriptive techniques to obtain the distribution for each segment. This technique was also used to identify significant factors towards health disruption, safety disruption, design of work schedule and individual element on shift working difficulty. Likert scale was used in the survey questionnaire form to measure workers’ perceptions on each question. Hayes [8] suggests the Likert scale based on five different categories, namely: (1) highly disagree, (2) disagree, (3) unsure, (4) agree, and (5) highly agree. For the purpose of this study, the authors had combined highly agree and agree and also highly disagree and disagree to two attributes, namely “agree” and “disagree” during data analysis process.

The survey questionnaires were distributed to Company X. The two main reasons for choosing Company X for the survey were due to their efficient management and human resource system. In this study, the survey questionnaires were distributed to 200 shift shop floor workers in the production department. The questionnaires were collected immediately to avoid missing questionnaire and most important, to ensure attention were given in answering the questions. The survey questionnaires distributed comprised three main sections. Section A covers the demographic information to describe the profiles of shift workers such as age, marital status and academic qualifications. The reasons to choose these criteria are inline with Rosa and Colligan [3] statement that states: age, gender, marital status, economy, individual factors and design of work schedule play important roles with respect to workers’ safety and health. Section B listed two main aspects with a number of factors or causes that are related to the objectives of the study; namely: health disruption, safety disruption. Section C had asked questions on the workers’ perception with respect to the shift system. In this study, aspects which received 60% or more points in term of agree and highly agree are considered as critical. Meanwhile, aspects which received 40% points in terms of disagree or highly disagree are considered uncritical.

4. Analysis Survey Results

4.1 Background of Company X

Company X was established on April 24th 1916 as a gold plate manufacturer in Klang Valley, Selangor, Malaysia. Later on, Company X progressed to become a manufacturer for producing sensors in automotive, aircraft and Heat Ventilation & Air Conditioning (HVAC) industries for local and international market. Company X produces four types of sensors namely; Micro-fused Gauge Sensors (MSG), Occupant Weight Sensors (OWS), Air classifications module (ACM), and Differential Pressure Sensors (DPS). The main customers of this company are Ford, Mitsubishi, Siemens and Honda. Company X employs about 1500 full-time staff, covering two systems of work schedule, normal shift and rotating shift. About 1000 shop floor workers, technicians, engineers, and others are working using the shift work system, while the rest (management division) practice normal work schedule. Company X had implemented a rotating 3 shift system that consists of 8 hours per shift, namely morning, evening and night shift.

4.2 Demographic Background

The shop floor workers comprise of various age groups: 40% less than 25 years old, 53.5% less than 46 years old, and 6.5% age above 46 years old. About two thirds of shop floor workers had married and more than 130 of them already have children. In term of monthly income, most of the shop floor workers are paid between Malaysian Ringgit (RM), RM 1001 to RM 2000. The percentage of shop floor workers’ income below RM1000 and above RM 2000 is approximately the same. The survey results shows about 77% of the shop floor workers had worked in Company X, 5 years or less. In term of academic qualification, 74% of the shop floor employees are educated up to the middle secondary school level. Furthermore, with respect to medical leave and attendance record, more than 60% of shop floor workers took medical leave and 43.5% of them were absent around 1 to 2 times per month.
4.3 HEALTH AND SAFETY FACTORS AND ITS CAUSES

There are two main aspects that could influence health and safety of shop floor workers at Company X, they are; health and safety effects. These factors are based on the previous studies on shift work that listed several elements which have influence on safety and health of shop floor workers. Under the previous four shifts system, employees at Company X work for five days, 40 hours each shift and take leave for 2 days after that, before changing to following shift. Each shift has a 40 minutes rest period. Based on an interview with the employees, the young employees like this shift system due to its potential to higher income. On the other hand, married employees are complaining with respect to the long working hours and insufficient time to rest, leisure and housework. In relation to the design of workplace, most of the employees at the shop floor are required to stand-up for 8 hours a day except for those who are working at Quality Control Department. In fact, stress and awkward posture caused by this working condition may have negative effects on their health and safety. From the survey questionnaires, the authors had listed down all the critical and uncritical aspects.

4.3.1 HEALTH EFFECTS

In the survey questionnaire, the authors had investigated fifteen issues on the health effects, which comprise of: (1) sleep deficit, (2) interrupted sleep, (3) stress, (4) fatigue, (5) depress symptoms, (6) headaches, (7) digestive disorder, (8) high blood pressure, (9) heart problem, (10) difficulty in managing daily life activities, (11) frequent fever, (12) lack of focus, (13) nutritional disorder, (14) lack of focus on family and friend, and (15) insufficient time for social activities. Results of the survey are shown in Figure 1.

![Figure 1: Health Effects among Shop Floor Employees](image)

With respect to Figure 1, cumulatively, 77% of shop floor workers had chosen “agree” to lack of focus with family and friends. In addition, 75.5% had responded “agree” on sleep deficit, 66% chosen both interrupted sleep and lack of time for social activities. As many as 65% shop floor workers reported to having difficulties in managing their daily life activities, 63% experienced fatigue, 57.5% have frequent headaches and 53% of them faced lack of focus as a result from shift work. On the other hand, of 55% shop floor workers were unsure whether working on night shifts have elevated themselves risks on developing heart problems and 48% were unsure about any perceived relationship between high blood pressure and working at night. Whereas, a few of effects are below 50%; 45.5% of shop floor workers faced depressed problem, 42% of shop floor workers faced mental and physical stress, nutritional disorder and fever; and 37% of shop floor workers experienced digestive disorder. From this research, it shows that there are six critical health disruptions found in the survey. The top two critical health disruptions showed large percentage gap from the rest; they are: 1) lack of focus with family and friends; 2) sleep deficit. Meanwhile, the other four health disruption had more than 60% responded “agree” on interrupted sleep, insufficient time for social activities, difficulty in managing their daily life and fatigue.
 Compared with people who work in normal working hours, shift workers report more interference to their family lives, especially the time available to spend with spouses and children. This fact is very important since the amount of time and quality of social interaction is related to physical and mental health. Individuals who cannot establish regular routines in their daily activities have difficulties planning for family responsibilities and coping with physical and mental fatigue as effectively as non-shift workers. This phenomenon is inline with Rosa and Colligan [3] statement, which stated that the morning shift employees on daily basis have to wake up from their sleep too early. As a result, it will cause sleep disruption and fatigue, thus reducing their quality of sleep and sleep time. In addition, they also faced problems in social interactions with their families and friends.

According to Costa [6], shift work can cause a negative effect to the human psychological condition. In fact, the human body is exposed to a continuous stress in its effort to regulate as swiftly as possible to the new working hours, while at the same time, it is being consistently frustrated by the continuous change-over. As a result of these problems, the employees will suffer anxiety, depression and digestion disorder. In the long run, they could result in gastrointestinal, neurophysic, and cardiovascular problems. In addition, a recent study by Tepas [9], found that due to the long shift hours, one must have strong physical and mental strength, especially for old employees.

In general, the survey results shown in Figure 1 are based on the short term effects due to the human body’s failure to accept and control the circadian rhythm. Approximately 50% of the surveyed shop floor workers were unsure whether they face a higher risk to heart attack and high blood pressure due to shift work. A more conclusive data could have been obtained if the shop floor workers were actually checked by a physician to determine their actual physical and mental health. One of the reasons why the shop floor workers were unsure of their actual health conditions are because they rarely undergo medical checkups, and they felt they are still young and healthy. This is reflected on the shop floor worker’s background, where more than 77.5% have been working for less than five years, and that 78.5% were less than 35 years old.

4.3.2 SAFETY EFFECTS

Figure 2 shows, six safety effects, namely: (17) lack of focus during work, (18) always make mistakes, (19) sleepiness during work, (20) fatigue, (21) involved in minor or major accident at workplace, and (22) involved in accident during traveling to work. With reference to Figure 2, cumulatively, 76.5% shop floor workers agreed that they feel sleepy during work, 58% experienced lack of focus, and 51.5% admitted that they make mistakes during working hours. On the other hand, there are three effects that registered less than 50%, namely; 46.5% experienced fatigue during work, 43.5% involved in minor or major accident at workplace, and 40.5% involved in accident during traveling to work.

Based on Figure 2, sleepiness during working time is the most critical safety disruption compared to another safety disruption. It is due to fatigue and insufficient sleep. It is related to the disturbances of circadian rhythms which can affect concentration, motivation, and reaction time, particularly at night. This combination can result in an increased risk of accident and injury. Besides, Tepas [9] also stated that instability sleep time will caused declining sleep quality and make the human employee feels sleepy especially during the afternoon. However, from this study, there is a disagreement about shift workers accident rates. It is not a very critical safety problem at this company. Only 40% of shop floor workers involved in accident while traveling and at workplace. Part of the reason is that work conditions are not always same on different shifts. For example, the amount of supervision, nature of workload and backup systems available can vary from one shift to another shift. Besides, the shop floor workers also agreed that adequate facilities and provided transportation help them a lot to avoid accident during travelling to workplace. According to Tepas [9], most accidents are caused by fatigue and they occurred from 12 mid night to 4 am in the morning. The factors that contributed to severe fatigue at this stage are insufficient rest period, sleep deficit, many physical activities, work surrounding, stress and lack of nutritional.
Section C had asked shop floor workers on their shift work schedule. This survey considers the length of rotation period, which consist of working day and rest period of at least 24 hours after each set of shift works. There are three questions, two objective questions and one subjective question. In the first question, comparison was made on average of sleep duration for the three shift systems. For the second question, it had asked on the type of comfortable shift system among the shop floor workers, which comprise of 5 working days and followed with 2 days leave, work for 4 days and take leave for 2 days after that, work for 4 days and take leave 3 days, and etc. The last question had asked about the improvement that needs to be done on their design of shift system. From the survey, most of the shop floor workers had suggested the improvement on rest period and also working conditions.

With reference to Figure 3, there are eleven items (labeled 23 to 33) that need to be considered in the design of work schedule: (23) number of consecutive work hours, (24) number of consecutive working days, (25) number of annual leave, (26) number of short breaks, (27) tired at work, (28) need for overtime, (29) need to work on public holidays, (30) frequent changes of schedule, (31) hard to do work rotation (morning, evening, night), (32) difficult to work on night shift, and (33) standing during working time.
In fact, it is very important to study the human factors which involved shop floor workers’ limitations, constraints, comforts and acceptance on job design. Figure 3 show about 81.5% shop floor workers complained that they are not comfortable with standing during working, 77% of shop floor workers said that they have insufficient rest duration, 74.5% of shop floor workers agreed that night shift is hard to track, and 68.5% of shop floor workers feels that their weekend leave is insufficient. Furthermore, 62.5% shop floor workers think that rotational shift is hard to follow, 57.5% of shop floor workers complained that there are too many working days, and 49.5% believed that their work need a lot of physical energy and they are forced to do overtime. On the other hand, about 52.5% of shop floor workers did not agree that they have to work during public holidays and 50.5% feel that work duration should not be extended.

Referring to Figure 3, standing during working time is the most critical factor at Company X. It is due to work requirements that require most of the shop floor workers to stand during work hours. In addition, more than 70% of shop floor workers agreed that short break and difficulty in working night shifts are two most critical disturbances factors at this company. There are two factors that had registered more than 60%, namely; insufficient holiday leave and difficulty in shift rotation. All the five factors are associated with the fatigue and other health disruption as discussed earlier. Tepas [10], pointed out that fatigue is related to two issues such as total time of each long shift and total of rest hours. Others stressed that hard physical work, continuous observation and insufficient rest may result in the deterioration of performance and reduce productivity among the employees.

4.5 SLEEP DURATION PER SHIFT AND SHIFT SCHEDULE

The shift work may alter the time and duration of sleep quality. The daytime sleep is seldom as deep and refreshing compared to sleeping at night. The problem is greater if the sleeping place is not a quiet, dark and comfortable. Even when disturbances are removed, a worker who returns home in the morning may still find it difficult to sleep or less refreshing. With respect to the sleep length, a reduced number of sleep hours’ is recorded for three shifts. Figure 4 shows the differences of sleep duration during night, morning and evening shift. The survey indicates that the employees experienced sleep deficit due to shift work, where 54% shop floor workers sleep in abnormal range (three to five hours a day). As a result, this problem which is known as circadian rhythm conflict will influence the employee’s productivity especially at night. Being constantly tired is a typical complaint of shift workers. In the long run, such condition may cause problems to the nervous system and in turn, increase the risk employees’ to health and safety.

![Figure 4: Sleep Duration Per Shift](image)

Optimizing the design of the shift schedule is the most effective way of reducing the health and safety problems. Satisfaction with a particular shift systems is the result of a complicated balancing act that is the best compromise for personal, psychological, social and medical concerns. Figure 5 shows the survey results on the most comfortable shift schedule as perceived by employees. Most of the shop floor workers are comfortable with working for 5 days with 2 days leave beginning the next day and work for 4 days with 3 days leave starting the next day. The percentage of both
schedule is approximately the same. From the survey data, the young employees had a tendency to choose work for 4 days because they can get extra leave. On the other hand, most of the old are employees comfortable to work for 5 days because their commitment to their families. Besides, working for 5 days need only 8 hours per day if compared to 4 days working day that need 12 hours per day. According to Tepas [9], the disadvantages of 12 hours per day are insufficient time to sleep, fatigue, and difficult to manage their daily life activities.

Figure 5: The Most Accommodating Shift Schedule Possible

5. CONCLUSIONS

Based on the survey data analysis, consideration for shift arrangements are associated with two main aspects, namely: work schedule and the behaviour of the employees. In addition, for health effects, lacks of focus with family and friends, and sleep deficit are the two major critical aspects that need to be considered by the company’s top management. Therefore, in the authors opinion the company should be conscious of the normal working period for their employees, 8 hours per day to help them plan their activity after work. This survey results is in-line with Neil [11], findings, which pointed out that the cooperation among employees and top management is needed to determine if the current design of the shift system reflect an overload or has negative influenced on their health and safety. Apart from that, sports championship and entertainment activity are others ways to bridge the relationship and promotes mental and physical fitness of the employees.

According to this survey on safety effects, 76.5% of shop floor workers are complaining that they feel sleepy during work and suggest the top management to be aware on this matter. Most of the shop floor workers agreed that by taking the stimulant or ‘pep’ drinks like coffee, cigarette and energy drinking, this sleepy feeling will go away. This behaviour will affect the health and safety of the employees in the long run. Therefore, it is very important that a shift worker involved in exercise programs to maintain an adequate level of fitness. It is also very important not to smoke, to have a good dietary habits and to participate leisure activities. From the study on design of work schedule, most employees agreed that short break and difficulty of night working shift are the two major factors which affect their health and safety condition. These factors are related to sleeping problem which faced by almost all shop floor workers. Apart from that, in term of safety, some of employees have been involved in accidents outside the workplace. To overcome this problem, the company had provided the transportation for its employees for coming and returning from work.

In this study, regardless of working time, standing posture while working can contribute to an accident. Awareness of existing work posture and practice is essential to identify changes that should be made to reduce the risk of accident. One must also adjust the work environment to meet each individual’s needs. The human body is very adaptive, and tries to accommodate itself to all situations. Poorly designed work spaces may lead to unhealthy postural adaptations. For that reason, the company should extend the break hour to one hour per shift for employees in order to avoid accidents at the workplace. For standing workers, the company can implement the work to break ratio of 30 minutes to 10 minutes, or 40 minutes to 10 minutes for each session. Providing anti-fatigue mat or chair for standing workers is an excellent example of ergonomic approach at this company. Standing on a hard surface is
uncomfortable because leg muscles become static, continuously flexed in an attempt to keep the body in an upright position. Additionally, providing adjustable worktable which corresponds to the size and employee’s height, can contribute to lessening effects of fatigue. A wide and flexible workstation design is also essential for the employees to change their position during working. By providing the foot rest, employees can transfer their weight to the feet. In addition, wrap-around support could stabilize the elbow and reduces pain caused by repetitive movements. In this day and age, manufacturing industry needs to operate 24 hours per day in order to fulfil the customers’ requirements and accomplish their expectations. Nevertheless, the survey findings show that the shift work can also present a negative impact on health, safety and well-being of workers. One explanation for this finding is that, it is well known fact that work efficiency during the night is not the same as during the day.

As a conclusion, this study was done successfully and had achieved its objectives to study on work schedule system in manufacturing industry and relate it to the health and safety of employees. It should be noted that the background among the shop floor workers differed quite substantially. Personal assessments of physical and mental well-being and personal characteristics such as gender, education level, marital status and age undoubtedly contribute to these aspects of this phenomenon. Human factors and ergonomics elements are the core elements of this study to evaluate and access the satisfaction level and restriction during shift work. Based on the findings from the survey, the authors had suggested the company to be more sensitive on the implication of shift working on shop floor workers towards their safety and health. The re-assessment of the work schedule is necessary for productivity improvement, quality, health and safety among the employees. The company also needs to consider the ergonomics design of the system and workplace surrounding, to achieve better health and safety among the employees during shift work.

7. ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Science Technology and Environment (MOSTE) Malaysia and Universiti Kebangsaan Malaysia for their support in providing the research grant for the project entitled “Establishment of Malaysian Anthropometrics Data for the Design and Development of safe and Comfortable Driver’s Seat for Road Vehicles” (Science Fund 03-01-02-SF0073).

REFERENCES

Health and safety perspective on medical waste management in a developing country: A case study of Dhaka city

M. A. Patwary a*, W. T. O’Hare a, G. Street a, K. M. Elahi b, S. S. Hossain c, M. H. Sarker a

a School of Science and Technology, University of Teesside, Middlesbrough, TS1 3BA, UK
b Environmental Science, Stamford University Bangladesh, Dhaka 1209, Bangladesh
c Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh

ABSTRACT

Mismanagement of medical waste may present an environmental health hazard. Causes of mismanagement include a lack of appropriate legislation, effective control, financial constraints and lack of education, training and awareness. This is specially a concern in developing countries like Bangladesh where a large number of people are involved in the formal and informal process of medical waste management. In this paper the management of the medical waste in Dhaka, the capital city of Bangladesh, was analyzed with an aim to propose an integrated medical waste management system in the city. Data was collected from Healthcare Establishments (HCEs) such as hospitals, clinics and diagnostic centers, waste disposal operatives, official and unofficial recycling operations and scavengers. Both quantitative and qualitative data gathering techniques were used, including observation and formal structured interview through questionnaire survey and informal dialogue.

The study focused on the character of HCE, its in-house and municipal management (collection, temporary storage, transportation, treatment and final disposal options of medical waste) practice in the study area. The results indicated that the management of medical waste had not been conducted properly. Deficiencies, inconsistencies and improper processes were revealed. Malpractice was observed, among individuals involved in handling medical waste, in mid-level management, and also in the senior management level. The study revealed the need for training, capacity building programs of all employees involved in medical waste management, to promote awareness and focus on environmental and health risks of medical waste. The results obtained may be generalised to most economically developing countries where there are similar environmental problems and strict budgets.

1. INTRODUCTION

The management of medical waste is an emerging worldwide concern. In the last few decades there has been a rapid growth of medical and patient related services around the world [1]. These have increased not only use of chemicals and drugs, but also increase in the use of disposable items [2] and consequently, large amounts of medical waste are being generated on a daily basis. As medical waste contains highly toxic chemicals and pathogens [3, 4, 5, 1, 6], it is generally considered a threat to environmental health and safety. It produces unpleasant smells and infestations. Infestations may in turn contribute to the transmission of diseases [8, 1, 9]. In most developing countries there is a growing concern that the enormous amount of hazardous medical waste generated, not only results in a huge disposal costs, but also creates the potential for the spread of diseases [7], Causes of medical waste mismanagement in developing countries include a lack of appropriate legislation, effective control, financial constrains, lack of education, training and awareness; and other related social problems [8, 9, 10].

In Bangladesh, like in many other developing countries, there is little emphasis on the proper management of medical waste [11]. Most of the advanced and comprehensive medical care and treatment services are available in Bangladesh. A number of reports have indicated that medical waste in Bangladesh has suffered from mismanagement practices that deviate substantially from the standard acceptable practices suggested by World Health Organization (WHO) [11, 13, 14, 15]. Such practices may create potential public health and safety risks, particularly to the individuals who are involved in waste related livelihood, but also to the general population.
The objective of this study was to investigate the contribution of mismanagement of medical waste to environment and health risk in Dhaka city. It was intended to gather information that would contribute to the development of medical waste management plans for developing countries, which could be adapted according to circumstances and which could be harmonised with the European Union Health-care Waste Management Regulations/Directives. It is likely that this would be helpful to minimize the risks of health and safety.

2. METHODOLOGY

2.1. SAMPLING AND DATA COLLECTION

Sampling: In this study the representative sample of HCEs was selected. Both purposive and authoritative sampling system were used to select individual respondents such as scavengers, waste collectors employed by Dhaka City Corporation (DCC), waste treatment facility employees, recycling operators and their employers, and related government and Non-Governmental Organization (NGO) officials and academics. Sampling methodology and sample size determination technique has previously been described [11, 12].

Qualitative Data: Qualitative data were obtained through observation, formal structured interview and informal interview as described below;

- Physical observation
  - HCEs (hospitals, clinics and pathology centers)
  - Medical waste treatment facility,
  - Medical waste recycling operations,
  - Government departments and NGOs related with medical waste activity
- Formal Structured Interview
  - Management authority of HCEs, waste treatment facility
- Informal interview
  - Relevant government, NGO and academic personnel,
  - Recycling operations’ employers.

Quantitative Data: Quantitative data was gathered through questionnaire survey among the individuals described below;

- Administration of questionnaire on
  - Employees of the various departments in the Health Care Establishments who are
    - Working directly with the patient care
    - Transferring waste from inside bins to road site bins
    - Working for mortuary departments.
  - DCC waste collectors, employed by DCC to collect waste from road side bins and to transport it to designated dumping places
  - Operators working at official medical waste treatment centers.
  - Scavengers involved in unofficial scavenging and resale of medical waste.
  - Individuals involved in informal recycling and repacking of medical waste.

2.2. DATA COLLECTION

The research methods (including the quantitative questionnaire and the structure of the formal interview) were approved by the Ethics Committee at Teesside University.

Initially, data was collected using an observational ethnographic approach [11, 12]. In the survey phase, administration of questionnaires was carried out between December 2005 and June 2006. Selected participants were briefed about the study, confidentiality was assured and written consent obtained. The questionnaire was pretested with a sub-sample of each group of selected participants.
Table 1 Methodological Approaches Adopted

<table>
<thead>
<tr>
<th>Category of information</th>
<th>NGOs personnel</th>
<th>Governmental personnel</th>
<th>Academic personnel</th>
<th>Employees of HCEs</th>
<th>Senior management of the HCEs</th>
<th>Waste treatment employees</th>
<th>Scavengers</th>
<th>Recycling operators</th>
<th>Recycling employer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Location, structure &amp; type of establishments</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Internal waste management</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>II</td>
</tr>
<tr>
<td>Waste transportation</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Recycling activity</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Hygiene, Health and safety</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Training</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>II</td>
</tr>
<tr>
<td>Legislation</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
<tr>
<td>Academic module</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
<td>FSI</td>
</tr>
</tbody>
</table>

FSI: Formal Structured Interview; II: Informal Interview

Table 1 shows how different methodological approaches (formal structured and informal interview) were used with each group to gather data under the various headings considered appropriate for that group. Responses were obtained from a total of 233 participants. The data were recorded and analyzed by descriptive statistical method using MINITAB (version 15).

3. RESULTS AND DISCUSSION

3.1. TYPE AND LOCATION OF THE HCEs

All of the clinics and pathology centers surveyed were private, while 50% of the hospitals were public and 50% private. The only medical waste treatment centre identified in the study was found to have been jointly established by an NGO, DCC and international donor agencies. It was operated by NGO and located on the outskirts of Dhaka city.

Table 2 shows that most of the HCEs, 41 (59.42%) were located at the main residential area, while 2 (2.89%) were located in industrial areas, 13 (18.84%) were located in commercial areas and 13 (18.84%) were located in mixed use areas (residential coexisting with government and private sector commercial activities). A few of the pathology centers (4%) were located at the industrial area. This pattern of location may present significant dangers in the context of urban population health and safety.

Table 2 Type and Location of Surveyed HCEs

<table>
<thead>
<tr>
<th>Type of location</th>
<th>Healthcare Establishments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>Hospitals (%)</td>
</tr>
<tr>
<td></td>
<td>2 (50%)</td>
</tr>
<tr>
<td>Industrial</td>
<td>0</td>
</tr>
</tbody>
</table>
Dhaka is a fast growing city, ranked 11th in the world by population (12.4 million) [19]. Expansion of the city is limited due to physical constraints. Commercial buildings are predominantly built and managed by the private sector, and mixed use developments, such as shopping malls and residential complexes, are popular. Most of the private HCEs are established in residential areas, or within shopping malls. Senior management and owners of the HCEs indicated, through informal interviews, that they have established their activities in residential areas to maximize convenience for their patients; as most people have no private transport, and there is no free ambulance service. HCEs, established in residential areas allow easy and timely access in an emergency. As there is a lack of legislation by the city authority, HCEs are allowed to dispose of their hazardous medical waste into general DCC waste containers intended for domestic waste. This leads to mixing of medical waste and domestic waste, resulting in all waste becoming hazardous. However, it was identified during the interview that management did not consider the health and safety issues for the residents of the risks of medical waste and some of them were not interested in discussing this issue.

### 3.2. **Type of establishments structure**

Of the surveyed HCEs, only 11 (15.94%) (including all 4 hospitals) were located in an enclosed site with more than one building while 35 (50.72) were housed in a single dedicated building, and 23 (33.33%) were located in a shared building (Table 3). All of the HCEs located in a single building or a shared in a building were privately owned. In some of those facilities examined, the space required for even basic provision for proper waste segregation, storage and disposal would significantly reduce the number of beds or tests that the facility could support and so reduce profitability. It was observed that none of the HCEs have an Effluent Treatment Plant (ETP) suitable for highly contaminated liquid waste. HCEs located in a single building discharged liquid waste directly to the municipal general sewerage system, and HCEs located on shared floors in a building discharged liquid waste into the building general sewerage. It was also found that some HCEs discharged liquid waste into adjacent lakes and rivers that were also used by local residents for washing and household purposes, as well as for agriculture. In 43 (62.31%) of HCEs liquid waste was discharged into a domestic septic tank system, and from there into the general sewerage system. The remaining 26 (37.68%) of HCEs were connected directly to the public sewage network. Wastewater was not being treated appropriately, the discharge may lead to potential contamination of drinking water supplies and/or environmental degradation.

During formal and informal interviews, respondents from management of these HCEs indicated that this too was due to lack of space, but also most of them did not have a positive attitude towards establishments of ETP. None of the HCE owners or managers was found to have considered waste management when establishing their business. Interviews with academics indicated that this was not due to the unwillingness of the architect or engineer, but due to the unwillingness of the HCE owner.

<table>
<thead>
<tr>
<th>Establishments Structure</th>
<th>Healthcare Establishments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hospitals (%)</td>
</tr>
<tr>
<td>Single building</td>
<td>0</td>
</tr>
<tr>
<td>Floor shared in a building</td>
<td>4 (24%)</td>
</tr>
<tr>
<td>Enclosed site with more than one building</td>
<td>5 (24%)</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
</tr>
</tbody>
</table>

### 3.3. **In-house waste management**

In general, from top to bottom management a lack of knowledge about the magnitude of the problem was found during interviews, and in most of the HCEs, there was no particular section or person assigned responsibility for the in-house management of medical wastes. Even where an individual was found to have been assigned responsibility for monitoring and managing waste, they were invariably assigned additional responsibilities and they considered waste management to be a minor aspect of their duties.
Hazardous medical wastes, should be collected, stored and transported separately in a designed bags or containers which are resistant to puncture. [8,17, 18]. The different colored bags having different types of wastes can be stored in the same compartment or in the same room. Most of the HCEs did not use designated bags or containers for the segregation and separate storage of medical waste. It was observed that most of the HCEs were using normal bags and containers which were not sufficiently robust. Once punctured, these allow hazardous waste to spread to the environment. On the other hand, in some instances, bags or containers designated with hazard signs and colored for hazardous waste were used for domestic waste as well. Sometimes waste bags and containers were placed around the HCEs reception area, emergency area, operation theatre, research laboratory or blood bank department.

Sharps waste should be crushed before placing them in a specially designated metal box [20]. However, in Dhaka sharps were not collected as required which may lead to injuries during the collection and transportation. In India, 85% of sharp injuries are caused between usage and disposal [8]. No respondents indicated that their HCE recorded any needle stick injuries, so data were not available. However, most of the waste workers (94%) had experienced accidental injury mostly from used needles and other sharps [12]. Of these, 28% were considered serious. This is much higher than the rates observed in other countries. In Japan 67.3% of HCE staff reported accidental injuries, compared with 50% in Peru and 18% in USA [14].

Collection and transportation in the surveyed HCEs was carried out primarily by HCE recruited waste collectors with little or no training. It was observed that sometimes they used patient trolleys to carry waste. Typically, the HCE workers did not wear sufficient protective equipment during waste handling activities, increasing the potential risk of accidents and personal injury.

### 3.4. Existing Waste Storage

Most of the surveyed HCEs did not have designated temporary storage facilities, again due to lack of space. Of the conducted HCEs 17 (24.63%) were operating internal temporary storage while 9 (13.04%) HCEs were operating external storage. It was observed that neither internal nor external temporary storage facilities were designated and designed for medical waste with adequate area and quality standards. None of the HCEs had refrigeration to store infectious pathological waste, as previously recommended [2, 18]. Only 5 (7.24%) HCEs were using large containers to store the bags of infectious waste. Usually, these containers were placed by the road outside the establishments where there was heavy pedestrian and car traffic. It was observed that the lids of the containers were mostly left open, so pedestrians also dumped all kinds of waste into the open containers. External storage, found in the surveyed HCEs, was not properly designed or identified. Sometimes open space within the HCE premises were used as an external storage. Crude storage leading to access to hazardous medical waste by birds, animals, flies and rodents may spread germs from contaminated waste to the surrounding community and environment. When internal spaces were used for storage, it was often observed that medical waste was placed in a room along with office stationery or medical supplies.

Data in Table 4 shows that 7 (10.14%) of HCEs were disinfecting storage facilities. Most of the HCEs were not using disinfectant for decontamination and were not installing appropriate equipment for final disposal of hazardous waste. Radioactive waste were not stored for a long enough period to decay and safely managed.

<table>
<thead>
<tr>
<th>Storage Structure</th>
<th>Hospitals</th>
<th>Clinics</th>
<th>Pathology/diagnostics</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporary storage</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>24.63%</td>
</tr>
<tr>
<td>Central or external storage</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>13.04%</td>
</tr>
<tr>
<td>Designated area for hazardous waste</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>13.04%</td>
</tr>
<tr>
<td>Disinfection for storage facility</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>10.14%</td>
</tr>
<tr>
<td>Adequate area and quality of the facility</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

### 3.5. Internal and External Transportation
Table 5 shows internal and external transport facilities used to carry medical waste within the HCEs and to waste dumping zone. None of the HCEs was found to be equipped with transport designated for disposal of medical waste within the premises. Therefore, waste workers transported medical waste by hand. Of the surveyed HCEs 5(7.24%) were found to be transporting medical waste from temporary storage to central or external storage area by using a hand held trolley with wheels which did not meet the regulatory requirements for safety (Table 5). Some of the trolleys had a closeable lid and some of them were without. The data obtained from the surveyed HCEs indicated that the time period for temporary storage was not followed strictly. Sometimes medical waste was stored for more than one day while the central storage period was in excess of two days.

Table 5: Internal and external transport facility of the conducted HCEs

<table>
<thead>
<tr>
<th>Transportation Structure</th>
<th>Healthcare Establishments</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport waste within the HCE by cart</td>
<td>Hospitals: 0</td>
<td>Clinics: 0</td>
</tr>
<tr>
<td>Transport from internal storage to external storage by hand held trolley</td>
<td>2*</td>
<td>3*</td>
</tr>
<tr>
<td>Transport facility to dumpsite by designated vehicle</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Transport non-hazardous and hazardous waste separately</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>HCE owned designated vehicle</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Locally made trolley operated manually

One NGO has established some medical waste management activity in Dhaka, primarily in the Dhanmondi area. Thus, in this area some HCEs (71%) were found to transport medical waste from the premises to the dumpsite separately and segregated as non-hazardous and hazardous through the operations of this NGO. In other areas the percentage of HCEs following these practices was much lower (10%). None of the HCEs was found to own a designated waste transportation vehicle. Consequently, the HCEs which were working with the NGO for medical waste management activity were fully dependent on the NGO transportation. While awaiting collection by NGO vehicle, the hazardous waste was stored insecurely, and sometimes for a long time, within the HCE premises. Thus scavengers were able to obtain access to the waste.

Waste from the remaining HCEs was transported by DCC waste crew to the general waste disposal site mixed with non-hazardous.

3.6 WASTE DISPOSAL AND RECYCLING

The survey results indicated that none of the HCEs were incinerating their infectious, pathological and sharps waste. Interviews with senior management and city officials suggest that this was due to a lack of incinerator and lack of interest by the city authority. An NGO had established a medical waste treatment and dumping facility. It was located close to houses and in a low lying area, connected with water bodies which may easily be contaminated by run-off and leachate containing pathogenic microbes. As previously mentioned, almost all HCEs discharged mixed domestic and medical liquid waste into sewer networks or into septic tanks.

None of the surveyed HCEs were currently operating any program to recycle general waste; such as office paper, cardboard, plastics, metal cans and selected glass. Most of the recyclable medical waste (RMW) was collected together with the general waste and disposed of into DCC bin or dumpsite along with other waste. Most of the HCE authorities were not aware about the RMW. Therefore, those RMW were sold mainly by the low class employee. Those employees have no idea about the risk to sold contaminated waste which may lead to significant threat more widely. It was observed and also found form the interview that the individuals who were involved with selling are more interested on income and easy access to get together in a large volume. Before selling the RMW they did not take any precaution measure if the waste was contaminated. It was found from the interview that waste was also sold by waste scavenger to recycle operator. None of the scavenger had any knowledge of risks from medical waste exposure. All of the recycle operators were found without any knowledge of risk from medical waste. Re-cycle operators were found to be involved with recycle activity without any training. It was observed and also found from the survey that re-cycle employers did not consider occupational health and safety for the employee.

4. TRAINING AND AWARENESS
The responsible personnel were not given adequate training, which leads to inappropriate management and insufficient implementation of the management system. Neither the HCEs nor the employees of the DCC in charge of collecting medical waste were trained. They were not aware of the importance of the appropriate handling and management of the medical waste, and consequently there were inappropriate activities observed during the collection and segregation of the waste generated from the HCEs.

Training programs on medical waste management for the medical staff (doctors, nurses and technicians) were limited. None of the HCEs offered any in-house training program. While some provided limited training for support staff through an NGO, this was the same program for all staff (senior management staff, administrative management, medical staff, waste worker, treatment facility staff). This supports the previous study of Ahmed et al. (2006) who found that 70% of HCEs did not offer any training, whereas Askarian et al. (2004) stated that training program and educational classes are instituted repeatedly for all personnel according to their job and the content of program is specifically designed in developed countries.

All of the waste workers indicated that orientation programs for new employees about medical waste management was not provided. Only 4 (5.79%) of surveyed HCEs displayed posters and some leaflets on hazards of medical waste to warn local residents of hazards, supplied by an NGO.

All of the HCEs and respondents (100%) indicated their needs and willingness to participate specialised training programs on medical waste management in future. Most of the HCEs 50 (72.46%) would prefer regular training, either quarterly or half yearly. All of the interviewed respondents (n=30) mentioned, training for senior management and authority was highly required while training programs for medical staff, waste handlers, and maintenance staffs were requested by more than 90% of the surveyed HCEs.

5. LEGISLATION

It was observed that none of the respondents were aware of legislative requirements about medical waste management. This was mostly because of the inefficient control and enforcement of the responsible authority. It was found from the survey that there was no legislation on medical waste management in Bangladesh. A set of regulations is being prepared by the Ministry of Environment and Forest, Government of Bangladesh which is still in the draft stage. However, authority could till apply to enforce for medical waste management in compliance with the Environmental ACT 1995 and Environmental Rules and Regulations 1997 of Bangladesh. It was found that the management of the medical waste in Bangladesh was supposed to be conducted according to the Environmental Rules and Regulations 1997. However, the Regulations were not applied by any of the surveyed HCEs.

6. RECOMMENDATION

The proposed efficient Medical Waste Management are given below:
- The waste should be categorised according to the WHO Directives and the European Waste Catalogue for the clear definition of hazardous medical waste and of its various categories in the Regulation.
- Location and structure of the HCE should be considered as a significant factor.
- Strict legislation of the medical waste creator regarding safe handling and disposal should be confirmed.
- Every stage of management should be documented.
- Autoclaving, sterilization, incineration must be considered before final disposal.
- If permit land filling should be considered as the final disposal methods. However, land filling would difficult in Bangladesh since it is a densely populated small country where it will be difficult to find appropriate land to fill. Therefore, waste generation reduction at source should be considered. Proper treatment and recline may minimize this problem.
- There should be provision in the legislation for the designation of courts responsible for handling any disputes may arise from the enforcement or of non-compliance with the law.
- A regular inspection system should be introduced to ensure that the enforcement of the legislation by a third party, penalties to be imposed for contravention and the inspection procedures should be clarified.

7. CONCLUSION

In this study, current medical waste management practices were investigated and evaluated in detail in Dhaka City, chosen as an example of the type of modern megacity found in developing countries. The findings suggest that there are serious problems in the safe handling and transportation of medical waste, probably arising from administrative and financial problems within HCEs, as well as the lack of awareness of the significance and the threats of the medical waste. These issues result in mixing of medical and non-hazardous waste, greatly increasing the total amount of hazardous waste. This could be addressed by promoting proper segregation through training,
clear standards and strict enforcement. These steps would also minimize the risk to those involved in handling waste, both within and outside the HCEs.

ACKNOWLEDGEMENTS

Authors acknowledge financial support from Charles Wallace Trust UK, Mountbatten Memorial Grants UK and The Hammond Trust, British Council for this research.

REFERENCES

High Efficiency Cloning of Drug Resistance Genes Via Safe Vector Host Systems

Connolly S.S.¹,²* and Connolly S.N.²

¹Department of Microbial Diseases, St John’s Institute of Dermatology, Guy’s, Kings College & St Thomas’s Hospitals, University of London, London, SE1 7EH, UK; present address: ²School of Science & Technology, University of Teesside, Borough Road, Middlesbrough, TS1 3BA, UK;

ABSTRACT

Direct cloning of the genes originating from clinical isolates of coagulase negative staphylococci (CNS) conferring high level resistance to mupirocin (Mup⁴, >2000 mgL⁻¹) is described using a vector system based on the safe non-pathogen Staphylococcus carnosus. This approach eliminates the need for initial laborious subcloning steps via S.aureus and E.coli shuttle vectors. Thus, an 11.5 kb fragment of a S.epidermidis plasmid conferring Mup⁴ was cloned into a variant of the pCE10 vector and introduced into S.carnosus strains by a modified protoplast transformation procedure, with high efficiency (8x10⁸) observed. In this way, straightforward cloning into a host able to express the gene without requiring genetic manipulation in a pathogenic strain curtails the wider dissemination of drug resistance elements. The opportunity for CNS gene expression in isogenic host species is afforded by this system. Evidence for the co-existence of elements conferring high level Mup⁴ both as a plasmid and an integrative form in the S.carnosus chromosome is also presented. Conjugal transfer of the hybrid plasmid to S.epidermidis also indicates successful cloning of the mobility (tras) genes as part of the Mup⁴ plasmid.

1. Introduction

Resistance to high (>512 mgL⁻¹) levels of mupirocin (Mup), conferred by a variety of plasmids, has been described in pathogenic and coagulase negative staphylococcal (CNS) species of human origin associated with clinical isolates [1] [2] [3] while both high and low level (<256 mgL⁻¹) resistance has been attributed to chromosomal resistant determinants [4] [5]. Large plasmids in the size range 23-68.4 kb have been reported in both S.aureus and CNS from various health care settings since the early 1990’s where the resistance predated the application of Mup [6]. Characterisation of these plasmids has demonstrated their different restriction profiles and ability to transfer to pathogenic strains by conjugation, particularly on the surface of skin.

The mupirocin resistance (Mup⁴) gene has previously been cloned as a 13.4 kb HindIII fragment of a S.aureus plasmid via a S.aureus-E.coli shuttle vector [7]. The present study involves an alternative strategy of direct cloning of the Mup⁴ gene from S.epidermidis via a 2.15 kb truncated S.carnosus vector based on pCE10 through modification of a protoplast transformation procedure with high efficiency. Thus although S.aureus and E.coli have been routinely employed as hosts, S.carnosus serves as a more amenable, non-hazardous and isogenic cloning and gene expression system due to its ease of transformation and non-pathogenic nature. The cloned Mup⁴ gene is stably maintained as part of a conjugative plasmid and also as an integral part of the chromosomal DNA in S.carnosus conferring high and low level resistance to Mup, respectively.

* Corresponding author: Tel.: (0044) 1642 342509; Fax: (0044) 1642 34201; E-mail: shahrzad.connolly@tees.ac.uk
2. MATERIALS AND METHODS

The Mup<sup>R</sup> plasmid was isolated from a clinical *S.epidermidis* strain, SCTD196, which also carried a penicillin resistance element, Pn<sup>R</sup>. Pn<sup>R</sup> was cured by sub-culturing SCTD196 and incubating at 40 °C. The characterisation of the Mup<sup>R</sup> plasmids from SCTD196 involved conjugative transfer of a 34 kb plasmid to a plasmid free strain, NCTC8325, which was chromosomally resistant to tetracycline. Serial conjugative transfer of the 34 kb replicon rendered various recipients resistant to Mup at a high level (>2000 mg L<sup>-1</sup>). Curing experiments involved culturing of SCTD196 and the Mup resistant recipient NCTC8325 at 40 °C in the absence of the respective antibiotics. Mup<sup>R</sup> plasmids were digested with *Eco*<sub>R</sub>I and *Hin*<sub>III</sub> (Sigma) according to the manufacturer’s instructions. Further characterisation of the *S.carnosus* Mup<sup>R</sup> clones involved hybridisation of the *Eco*<sub>R</sub>I digested Mup<sup>R</sup> plasmids with a 4 kb as well as a 13.4 kb labeled Mup<sup>R</sup> DNA probe [1]. Bacterial strains and plasmids are shown in Table 1.

![Table 1: Characteristics of the strains employed](image)

<table>
<thead>
<tr>
<th>Strain</th>
<th>Species</th>
<th>Characteristics</th>
<th>Resistance Determinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCTD196</td>
<td><em>S.epidermidis</em></td>
<td>Clinical isolate</td>
<td>Mup, Pn</td>
</tr>
<tr>
<td>SCTD196C1</td>
<td><em>S.epidermidis</em></td>
<td>Cured of Pn&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mup</td>
</tr>
<tr>
<td>SCTD196C2</td>
<td><em>S.epidermidis</em></td>
<td>Cured of Mup&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Sensitive</td>
</tr>
<tr>
<td>TM300-pCA43</td>
<td><em>S.carnosus</em></td>
<td>Host and vector plasmid pCA43</td>
<td>Cm, Asa, Asi, Sb</td>
</tr>
<tr>
<td>SA113-pCE10</td>
<td><em>S.aureus</em></td>
<td>Host and vector plasmid pCE10</td>
<td>Em, Cm</td>
</tr>
<tr>
<td>SA113-pCE10V</td>
<td><em>S.aureus</em></td>
<td>Host and vector plasmid pCE10 truncated</td>
<td>Em</td>
</tr>
<tr>
<td>STH7C</td>
<td><em>S.aureus</em></td>
<td>Transconjugant originated from NCTC8325 mutated to Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Tc, Mup</td>
</tr>
</tbody>
</table>

Mup, mupirocin; Pn, penicillin; Cm, chloramphenicol; Asa, arsenate; Asi, arsenite; Sb, antimony; Em, erythromycin; Tc, tetracycline

An 11.5 kb *HindIII* fragment from the plasmid SCTD196 was inserted into a *S. carnosus* host strain via a modified 2.15 kb pCE10V plasmid. The same insert fragment was also cloned onto a 5.9 kb *S.carnosus* vector pCA43. Hybrid plasmids were introduced into *S.carnosus*, TM300, by a modification of the previously reported method of protoplast transformation using plasmid DNA concentrations in the range 250-2500 mg L<sup>-1</sup> [7]. The protoplast suspension was incubated at 26-28 °C for 2 min with the DNA solution preheated to the same temperature prior to its addition. The mixture was cooled to room temperature and centrifuged at 7000 rpm for 25 min. Protoplast dilutions were inoculated on DM3 regeneration medium and incubated at 37 °C for 5 hours before the addition of 3 mL of softagar containing appropriate antibiotics. Plates were incubated for 3-4 days. Transformants were selected on Mup and erythromycin, Em, solid media, and confirmed according to restriction analysis and hybridisation to a Mup<sup>R</sup> DNA probe.

Plasmid DNA was isolated according to alkaline lysis with the addition of 5.2 mg L<sup>-1</sup> lysostaphin (Sigma); purification was effected according to a standard phenol-chloroform method [8]. Electrophoresis was conducted on 0.8-1.2% agarose gels using Tris-borate-EDTA (Sigma). T4 ligase, DNA labeling and detection kits, as well as other reagents were obtained from Sigma. Other strains were routinely cultivated on Oxoid blood agar base containing Cm, 15 mg L<sup>-1</sup>; Em, 15 mg L<sup>-1</sup>; and Mup, 100-200 mg L<sup>-1</sup>.

3. RESULTS

Acquisition of the 34 kb plasmid by the recipient strains through conjugation and protoplast transformation conferred high level Mup<sup>R</sup>. The resistance was stably maintained and serially transferable to new *S.aureus* strains by conjugation at the rate of 2 x 10<sup>8</sup> to 3 x 10<sup>7</sup>. Loss of the plasmid under curing conditions rendered the strains sensitive to Mup. The efficiency of transformation of *S.carnosus* in molecular cloning of the Mup<sup>R</sup> gene and protoplast transformation was 4.8 x 10<sup>5</sup> to 8 x 10<sup>4</sup> for concentrations of hybrid DNA up to 500 mg L<sup>-1</sup>. Significantly higher efficiencies of transformation in the order of 2 x 10<sup>5</sup> to 8 x 10<sup>5</sup> were obtained for DNA concentrations of between 1000-2500 mg L<sup>-1</sup>. Characterisation of the plasmids using *HindIII* and *Eco*<sub>R</sub>I demonstrated the presence of the same fragments from the Mup<sup>R</sup> plasmid in both the clinical strain and the transconjugants; see Table 2 for the restriction profiles of the plasmids used in cloning.
while the strains remained resistant to 5.00 mgL\(^{-1}\). Curing of the Mup resistant Staphylococcus strains was achieved by treatment with a linearised plasmid (pSSC40) which was used to transform the recipient strain. The transconjugants and the clinical strain hybridised to the Mup\(^R\) probes confirming the presence of the Mup\(^R\) gene in all resistant strains. The hybrid plasmid pSSC40 recombinant generated through cloning into Staphylococcus carnosus was isolated from clones of Staphylococcus carnosus and cleaved with HindIII resulting in fragments of 11.5 and 2.15 kb. These were separated by gel electrophoresis and transferred to a nylon membrane. The 11.5 kb HindIII fragment of the Mup\(^R\) plasmid from Staphylococcus carnosus clones was hybridised to labeled 13.4 kb and 4 kb Mup\(^R\) gene probes, thus ascertaining the presence of the Mup\(^R\) gene in Staphylococcus carnosus clones. These probes were also used to detect homologous sequences in the plasmids of transconjugants and the clinical strain. The 11.5 kb fragment from the transconjugants and the clinical strain hybridised to the Mup probes confirming the presence of the Mup\(^R\) gene in all resistant strains.

Curing of the Mup resistant Staphylococcus carnosus clones led to the complete loss of the extrachromosomal 34 kb plasmid, while the strains remained resistant to 500 mgL\(^{-1}\) of the antibiotic. A chromosomal DNA fragment of 10 kb in Mup resistant Staphylococcus carnosus clones hybridised to the Mup\(^R\) probe indicating the presence of a copy of the Mup\(^R\) gene within the chromosome. The chromosomal Mup\(^R\) determinant was stably maintained under curing conditions and was not transferable by transconjugation.

### Table 2: Product of the restriction enzyme cleavage of plasmids

<table>
<thead>
<tr>
<th>Strain source of plasmid</th>
<th>Restriction enzyme</th>
<th>Mean restriction fragment size (kb)</th>
<th>Size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCTD196C1</td>
<td>HindIII</td>
<td>11.5, 7.8, 5.2, 4.4, 2.9, 2.4</td>
<td>34.2</td>
</tr>
<tr>
<td>SCTD196C1</td>
<td>EcoRI</td>
<td>10.8, 9.7, 5.4, 3.9, 2.6, 1.8</td>
<td>34.2</td>
</tr>
<tr>
<td>TM300-pCA43</td>
<td>HindIII</td>
<td>linearised</td>
<td>5.9</td>
</tr>
<tr>
<td>SA113-pCE10</td>
<td>HindIII</td>
<td>3.05, 2.15, 0.1</td>
<td>5.3</td>
</tr>
<tr>
<td>SA113-pCE10V</td>
<td>HindIII</td>
<td>linearised</td>
<td>2.15</td>
</tr>
<tr>
<td>SCTH7C</td>
<td>EcoRI</td>
<td>10.8, 9.7, 5.4, 3.9, 2.6, 1.8</td>
<td>34.2</td>
</tr>
<tr>
<td><em>S. carnosus</em> pSSC40</td>
<td>HindIII</td>
<td>11.5, 2.15</td>
<td>13.65</td>
</tr>
</tbody>
</table>

*a* pSSC40 recombinant generated through cloning into *S. carnosus*

The hybrid plasmid pSSC40 was isolated from clones of *S. carnosus* and cleaved with HindIII resulting in fragments of 11.5 and 2.15 kb. These were separated by gel electrophoresis and transferred to a nylon membrane. The 11.5 kb HindIII fragment of the Mup\(^R\) plasmid from *S. carnosus* clones was hybridised to labeled 13.4 kb and 4 kb Mup\(^R\) gene probes, thus ascertaining the presence of the Mup\(^R\) gene in *S. carnosus* clones. These probes were also used to detect homologous sequences in the plasmids of transconjugants and the clinical strain. The 11.5 kb fragment from the transconjugants and the clinical strain hybridised to the Mup probes confirming the presence of the Mup\(^R\) gene in all resistant strains.

Curing of the Mup resistant *S. carnosus* clones led to the complete loss of the extrachromosomal 34 kb plasmid, while the strains remained resistant to 500 mgL\(^{-1}\) of the antibiotic. A chromosomal DNA fragment of 10 kb in Mup resistant *S. carnosus* clones hybridised to the Mup\(^R\) probe indicating the presence of a copy of the Mup\(^R\) gene within the chromosome. The chromosomal Mup\(^R\) determinant was stably maintained under curing conditions and was not transferable by transconjugation.

### 4. DISCUSSION

Molecular cloning and the expression of staphylococcal genes has been achieved via *S. aureus* and *E. coli* plasmids as shuttle vectors for the transformation of *S. carnosus* protoplasts [7] [9]. Other approaches, including electroporation for the insertion of exogenous DNA into cells and transcutaneous electrical nerve stimulation for horizontal gene transfer from CNS, have also been used with various frequencies of transformation involving *S. aureus* as a recipient [10] [11] [12] [13]. The limitations presented by handling hazardous pathogenic strains, including *S. aureus*, and the non-isogenic character of both *S. aureus* and *E. coli* have led to the promotion of *S. carnosus* as the preferred host system for expressing staphylococcal genes. The major advantage of *S. carnosus* is that it provides a safe host for gene expression due to its non-pathogenic nature. Recognition of the inert properties of *S. carnosus* and the non-isogenic character of both *S. aureus* and *E. coli* have led to the promotion of *S. carnosus* as the preferred host system for expressing staphylococcal genes. The major advantage of *S. carnosus* is that it provides a safe host for gene expression due to its non-pathogenic nature. Recognition of the inert properties of *S. carnosus* and the non-isogenic character of both *S. aureus* and *E. coli* have led to the promotion of *S. carnosus* as the preferred host system for expressing staphylococcal genes.

The introduction of hybrid DNA into *S. carnosus* protoplasts has generally relied on small concentrations of transforming open-circular DNA of low molecular weight resulting in low efficiency transformation [15] [7]. Significantly improved efficiency of transformation of the order of 8x10\(^9\) has been achieved by application of DNA at higher concentrations (1000-2500 mgL\(^{-1}\)). Other modifications include preheating the DNA suspension to 26-28°C before its addition to the protoplast suspension maintained at the same temperature; this is in contrast to previous reports of protoplast transformation and electroporation [15] [16] [7] [14].

The cloned Mup\(^R\) gene is stably maintained in *S. carnosus* as a 13.65 kb plasmid as well as an integral element within the chromosomal DNA. High level Mup\(^R\) (>2000 mgL\(^{-1}\)) is associated with the gene located on the plasmid, whereas resistance at lower levels (<500 mgL\(^{-1}\)) is attributed to chromosomal determinants. Curing and conjugal transfer experiments demonstrate that the 13.65 kb plasmid encodes Mup\(^R\) and carries the essential elements for its own mobilisation. High level Mup resistance was lost during curing with concomitant loss of the 13.65 kb element, while the *S. carnosus* strains remained resistant to lower levels of Mup. This suggests the presence of a stable chromosomal location for the resistance genes which was further confirmed by hybridisation of the EcoRI digests of chromosomal DNA with a Mup\(^R\) probe. Chromosomally resistant *S. carnosus* strains which were cured of their plasmid (pSSC40) did not transfer Mup resistance through conjugation with *S. epidermidis* and *S. aureus*, in contrast to *S. carnosus* strains which participated in serial transconjugation.
The presence of chromosomal determinants encoding both low and high level Mup resistance in *S. aureus* has been reported [17] [5], however, high level resistance has been mainly attributed to the presence of various conjugative plasmids in *S. aureus* and CNS [1] [18] [4]. In contrast, the present study demonstrates that the Mup<sup>R</sup> element is maintained in *S. carnosus* as both a plasmid and a chromosomal determinant. The protoplast transformation procedure described here serves as a versatile and efficient system for the cloning and expression of staphylococcal genes within an isogenic, apathogenic host.

**REFERENCES**


Introducing a lean recruitment framework for manufacturing

Paul M Gibbons
EngD Centre in Systems, University of Bristol, Bristol, BS8 1TH, UK

ABSTRACT

This paper introduces a new approach challenging the current paradigm for the recruitment processes employed by manufacturing organizations. Typically current recruitment methods focus on a convergent process through the short listing of candidates validated against the submission of a curriculum vitae or job application form before assessment against a standard company interview template. Hiring managers work closely with their human resources subject matter experts looking for the most effective method of assessing candidate suitability to perform to the job specification requirements balanced against their own constraints such as time and availability. Problems with this type of recruitment process can lead to a subjective assessment of the candidate’s suitability to match the real requirements of the hiring manager which are only identified once the candidate has been employed. The new recruitment framework draws from two disparate theoretical frameworks homogenizing them into a new and holistic recruitment process. Drawing from the lean production framework a focus on the value adding elements in the eyes of the customer – in this case the hiring manager – identifies waste in the current recruitment process. Similarities identified in a new product introduction, total design conceptual framework are assimilated into the new process transforming the product design specification tool into a position profile specification outlining the hiring manager’s requirements. A future state map of the hiring manager’s requirement is realized by combining the position profile with a Pugh matrix tool identifying potential candidate strengths, weaknesses and similarities to the role requirements. As with the total design framework, where the wrong choice of concept can rarely be recouped by brilliant detailed design work; the wrong choice of candidate at the recruitment stage cannot necessarily be recouped by brilliant personnel management skills when the selected candidate is employed. The lean recruitment framework is introduced as an objective recruitment framework focusing on the value adding steps to ensuring the right candidate is employed matched to the realistic requirements of the hiring manager.

1. INTRODUCTION

Typically, recruitment processes focus around the short listing of candidates validating suitability against the submission of a curriculum vitae or job application form before further validation against a standard company interview template through a multiple interview process. Divergent information gathered on individual candidate suitability is compiled in parallel to a convergent decision making process identifying the most suitable candidate overall. Hiring managers work closely with their human resources subject matter experts looking for the most effective method of assessing candidate suitability to perform to the job specification requirements balanced against their own constraints such as time and availability. Observed problems with this type of recruitment process can lead to the subjective assessment of a candidate’s suitability to match the real requirements of the hiring manager which are only identified once the candidate has been employed.

This paper argues a more holistic approach - taking into account the manufacturing strategy whilst encompassing a focus on the functional requirements of the role - will deliver a lean and more effective recruitment process. Adopting an action research approach [1], as have other lean researchers [2-8], the framework development is embedded within the researching organisation. Testing the lean recruitment framework in parallel to the extant recruitment process, the Pugh concept matrix [9, 10] is introduced as a useful and transferable tool. Developed from its origins as an engineering concept design review tool, the Pugh concept matrix is further enhanced into an iterative decision making tool for identifying candidate strengths, weaknesses and similarities against a datum point of individual position profile criteria.

Outputs of the research include a proposed lean recruitment framework introducing a data based candidate selection process focusing on the value adding elements through the recruitment process.
2. LEAN PRODUCTION

Reference [11] are responsible for the seminal work introducing the lean production concept as it is know today suggesting a focus on waste elimination of individual processes at what could be considered the micro level of processes employed within a wider operation. The conceptual development was based on the findings from a 5-year international motor vehicle project (IMVP) [12] based at the Massachusetts Institute of Technology (MIT). Creating mechanisms for industry-government-university interactions at an international level, the project team set six criteria for which its success would depend on: thoroughness; expertise; a global outlook; independence; industry access; and continuous feedback. To achieve the criteria, specialists from the automotive industry were sourced from around the world to carry out case study investigations within the global automotive industry. A key finding from the study suggested the lean production concept combines the best elements of craft production with those of mass production offering a pluralist approach which focuses on delivering:

“...reductions in costs per unit and dramatically improving quality while at the same time providing an ever wider range of products and ever more challenging work”.

Overall, the study revealed the existence of a 2:1 productivity difference between car assembly plants in Japan and those in the West. These findings lead to extensive industry soul-searching [13] which resulted in further benchmarking studies confirming the lean production project revelations [14-17].

In 1994 the tangential concept of a lean enterprise [5] was introduced as a development of lean production taking a broader perspective of the extant relationships between disparate processes within the overall system of manufacture; at what could be considered a more macro level of the overall operation. The ideas behind the lean enterprise were further complemented by a set of principles presented as lean thinking [6, 18] summarised as the 5 steps to becoming lean:

1. Specify value
2. Identify the value stream
3. Make the value creating steps flow
4. Promote a “pull” culture
5. Pursue perfection

Concurrent with the fundamental conceptual work [11] was the dissemination of the lean concept to other areas of operations management research. This research culminated in the lean supply [19, 20] and lean logistics [21] concepts. Other simultaneous and tangential work led to the development of new tools to enhance the original concepts including: seven value stream mapping tools, [2, 3]; value stream mapping [22] & big picture mapping [23]. Typically value stream mapping has been used as a static analysis tool quantify the 7 wastes of lean [24] defined as:

1. Overproduction
2. Waiting
3. Transport
4. Inappropriate processing
5. Unnecessary inventory
6. Unnecessary motion
7. Defects

The most recent phase in the evolution of lean has seen the introduction of the lean consumption concept [7, 8] which is described as “a necessary and inevitable complement to lean production and is about ... providing the full value consumers desire from their goods and service, with the greatest efficiency and least pain. Similarly to other dimensions of the lean concept [19-21] the principles correspond closely with the 5 steps to becoming lean [6, 18]. This more holistic approach encompasses new ideas and in particular is interested in supplying what the end customer wants, where they want it, when they want it and without wasting their time at any time; an end customer satisfaction focused approach to the lean concept.
3. PUGH MATRIX CONCEPT

Reference [9] is responsible for the seminal work introducing the concept of total design, a systematic activity focused on satisfying the needs of the customer through a design process encompassing product, process, people and organization attributes. Reference [9, 10] developed tools focused on the product attribute development arguing this is the most important element of total design as businesses cannot exist without products. Developing the framework further, product development in the total design process is based around the needs of the customer articulated through a product design specification (PDS). The PDS is used as a datum point through the design process placing a boundary around potential design concepts matched to the customer requirements. Critical to the successful development of a new product design is the ability to generate solutions to the design problem and then evaluate these solutions selecting the one which is best suited to the PDS. Reference [9] argues in practice, the wrong choice of concept can rarely be recouped by brilliant detail design. Facilitating the design decision process a comparison matrix is introduced as a useful evaluation tool for comparing different concepts against individual design criteria identified in the PDS.

Figure 1 presents the Pugh concept evaluation matrix model [25] as introduced in the total design process [9]. Operationalising the matrix, each design concept is compared to a pre-defined datum point of design criteria (as defined in the PDS). Decisions are then made by the design team as to whether the individual concept elements are the same as the datum point (indicated by an ‘S’), better than the datum point (indicated by a ‘+’), or finally worse than the datum point (indicated by a ‘-’). Once the matrix has been fully populated a summary table at the bottom of the matrix shows the individual concept’s ability to match the requirement as well as the strengths and weaknesses in the concept design.

<table>
<thead>
<tr>
<th>Concepts</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>……</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>S</td>
<td>-</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>T</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>U</td>
<td>+</td>
<td>+</td>
<td>S</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>n</td>
<td>S</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 1: Pugh concept matrix

Reference [9] suggests key outputs of the evaluation matrix include:
- A greater insight into the requirements of the specification
- A greater understanding of the design problems
- A greater understanding of the potential solutions
- A knowledge of the reasons why one concept is stronger or weaker than another.

4. RECRUITMENT PROCESS: CURRENT STATE MAP

The key principle supporting the lean conceptual framework has been identified as a need to understand what is value adding from the customer’s perspective [6, 18] and any subsequent non-value adding activities are classified as waste and can be eliminated [24]. For the recruitment process the customer would typically be the hiring manager
who will usually look for support from the human resource subject matter experts taking reference to company recruitment policies and procedures. Therefore, key to the identification of a lean recruitment process is an understanding of what is value adding in the recruitment process for the hiring manager and for the business overall. Once the value adding element is understood -using the value stream mapping approach from the lean conceptual framework- a current state map can be completed as a platform to build a future state map of the required process which can be achieved using staged plans known as loops. The use of staged plans is sometimes necessary as in some cases the realization of a future state map in one big leap is not always possible and therefore the ideal state map can be used as an end goal pseudo objective utilizing future state maps as incremental leaps.

Figure 2 presents an input-process-output model of a typical recruitment process in this case identifying the current state of the hiring manager’s value stream. Key inputs focus on the business strategy and needs giving a requirement to source a suitable candidate through the recruitment process culminating in the final selection and subsequent hiring. The value adding elements of the identified recruitment inputs-process-outputs would focus on the key outputs including the selection of a suitable candidate to match the requirements of the hiring manager and the supporting documentation compiled during the recruitment process which could be used by the hiring manager to plan any personal development required for the recruited candidate.

Reviewing the key inputs to the identified process there is an overlap of information relating to the documentation prepared by the hiring manager including job description/position profile, internal/external job advert and interview template. Matched to this is the submission by potential candidates of completed job application forms, CVs or both. Through the recruitment process, reference is made to the candidate's job application form and/or CV and further data is collected during the interview phase as to the candidates’ suitability to match the requirements of the hiring manager using the standard company interview template.

5. RECRUITMENT PROCESS: FUTURE STATE MAP & THE LEAN RECRUITMENT PROCESS

In the eyes of the customer -in this case the hiring manager- the value adding element of the recruitment process has been defined as the final selection of a suitable candidate matched to the real requirements of the business realized through the deployed manufacturing strategy. The current state map of the recruitment process has identified process waste extant in the form of a duplication of customer owned reference documentation, multiple comparison documentation submitted by the potential candidate and added to through the short listing and interview phases by the customer and finally a lack of focus through the selection process on the real requirements of the hiring manager resulting in a subjective decision making process based on divergent and disparate data.

The future state map should therefore look to focus on the requirements of the hiring manager using these requirements as a central datum point during the recruitment process. Subsequently, any reference material and investigative techniques should focus on the comparison of the candidates suitability matched to the requirements of the hiring manager as defined in the position profile/job description owned by the hiring manager. Reference to this documentation through the recruitment process will ensure the final selection is based on the specific requirements of the hiring manager ensuring the selected candidate will fulfill their own as well as the business expectations.
Operationalising the future state map, key inputs to the recruitment process should focus on the preparation of a position profile by the hiring manager matched to their manufacturing strategy. This document can then be used to create an internal/external vacancy advert before being used throughout the recruitment process as a central datum point to compare candidate suitability. Also important is to ensure candidate applications are made using a standard application form therefore collecting data which is easily transferable in comparison to the position profile. This would be especially useful at the short listing element of the recruitment process where the hiring manager will have many applicants to consider matched to their own time constraints.

Revisiting the total design conceptual framework focusing on the product design specification (PDS) [9, 10] and the Pugh concept comparison matrix [25], there is opportunity to adapt and transfer this approach to the recruitment process forming a foundation for the creation of a lean recruitment conceptual framework and more importantly a method of realizing the future state map requirements. Developing the lean recruitment conceptual framework, the PDS in the design phase can be used to create a datum point of key criteria to the overall requirements of the designer. Similarly, the position profile suggested in the future state map can be populated with key criteria to the overall requirements of the hiring manager. Critical to the lean recruitment process at this stage is the capture of realistic requirements of the hiring manager as specific criteria in the position profile, as this will be the datum point for candidate comparison through the recruitment process. Also important is the matching of the position profile criteria to the manufacturing strategy ensuring alignment to both business and departmental objectives. During the candidate selection phase the hiring manager can compare individual candidates (the concept designs) to the requirements detailed in the position profile (the PDS).

The second element of the lean recruitment conceptual framework is the introduction of an objective decision making process when comparing candidate suitability. The framework argues the use of the Pugh matrix tool allows for the comparison of individual candidates matched to a datum point of key criteria illustrated within the position profile. Figure 3 presents a conceptual template of the lean recruitment Pugh matrix using possible criteria examples typically found in position profiles which can be added to as defined in the position profile. The matrix is populated by comparing potential candidates to an antecedent set of functionally required datum points. Each candidate is compared against the individual position profile criteria datum point to see if they are better than the requirement (marked as a plus sign '+' and seen as a potential strength), or the same as the requirement (marked as an 'S' and seen as matching the requirement exactly) or finally, whether they are worse than the requirement (marked as a minus sign '-' and seen as a potential weakness). At the short listing stage, rather than complete the whole matrix for each potential candidate, the hiring manager can focus on comparing candidate suitability against a sample of key criteria. As the number of potential candidates is reduced through the recruitment process, more of the position profile criteria can be incorporated into the Pugh matrix facilitating convergent decision making through controlled divergent data collection.

Once the matrix is fully populated the summary of strengths, weaknesses and similarities is tabulated to show the potential candidates’ identified strengths, weaknesses and similarities to the overall requirement. The matrix can also be developed further by weighting significant criteria as well as adding other important selection criteria not detailed in the position profile such as candidate availability if the requirement is urgent. Once the matrix is fully populated the final stage of the comparison process is to rank the potential candidates in order of suitability. At this point potential candidate weaknesses can be noted for further investigation at the next stage of the recruitment process or, if the candidate is selected for the role, the weaknesses can be carried over to their personal development plan as training requirements. Also at this point, the potential candidate strengths can be noted for further investigation at the next stage of the recruitment process. If the potential candidate has scored very highly this may indicate they are over qualified/experienced for the role and therefore may not be suitable. However, the potential strengths identified may be useful for developing other employees and therefore the hiring manager should incorporate this into the decision making process.

As part of the recruitment process the Pugh recruitment matrix decision criteria is populated at the time of the hiring manager developing a position profile and is then updated throughout the recruitment process from short listing through to 1st interview, 2nd interview and hiring. During the recruitment process the hiring manager accumulates a portfolio of candidate suitability summarized in the lean recruitment Pugh matrix. For the candidates not selected their portfolio can be archived for future reference if required and the selected candidate’s portfolio can be carried over to their personnel file and to the hiring manager’s departmental skills matrix.

In summary, the future state map and lean recruitment framework eliminate the extant duplication of documentation by using the position profile (PDS) as a standard document through the initiation, advertising, selection and post hiring phases of the recruitment process in parallel to using a Pugh recruitment matrix to control the convergent candidate selection in a balanced divergent candidate data collection process.
6. SUMMARY & FINDINGS

The lean recruitment conceptual framework was developed and tested using an action research approach [1] using four sequential pilot studies combining case study research [26] with an action research process of planning, observing and reflecting [27] summarized as taking an action case research design [28]. Reflecting on the usefulness of the lean recruitment framework the benefits are very similar to those identified within the total design approach to product development [9, 10]:

- **A greater insight into the requirements of the specification (in this case the position profile).** Through the recruitment process the hiring manager gets the opportunity to question why the requirement is important to the position profile and overall manufacturing strategy. Taking an iterative approach, the requirement criteria can be changed to ensure alignment with departmental objectives through the convergent and controlled divergent lean recruitment process.

- **A greater understanding of the design problems.** Through the selection process the hiring manager gains an understanding of how realistic the requirement is in comparison to candidate availability. This could mean there is a skills shortage in that particular field of expertise or that the hiring manager has set an unachievable requirement. In summary, the hiring manager will gain an understanding of candidate availability to best match the requirement.

- **A greater understanding of the potential solutions.** Linking to the understanding of design problems the hiring manager -through the Pugh lean recruitment matrix summary- gets an understanding of the possible candidates who could fulfill the roll most adequately. In particular reviewing any potential weaknesses that would require further development or any candidate strengths that could be used in the development of other personnel.

- **A knowledge of the reasons why one concept is stronger or weaker than another.** Potential candidate suitability, strengths and weaknesses are investigated through a controlled divergent process in parallel to validation through the interview phase converging at the final selection decision stage. Through the lean recruitment process, each candidate is compared against the individual position profile criteria.
datum point to see if they are better than the requirement (marked as a plus sign ‘+’ and seen as a potential strength), or the same as the requirement (marked as an ‘S’ and seen as matching the requirement exactly) or finally, whether they are worse than the requirement (marked as a minus sign ‘-’ and seen as a potential weakness). Once the final lean recruitment matrix is fully populated the summary of strengths, weaknesses and similarities is tabulated to show the overall candidates’ potential to meet the requirement. The summary table can also be ranked to show order of candidate suitability and key criteria can be weighted to ensure the candidate assessment is correctly balanced to the requirements.

In summary the lean recruitment framework eliminates the duplication of documentation, gives ownership of the recruitment process to the hiring manager and provides a convergent/controlled divergent structured approach focusing on the realistic requirements of the hiring manager through the recruitment process.

7. CONCLUSIONS

The main benefit of the lean recruitment process is the shift in emphasis to focus on the value adding elements of the recruitment process in parallel to placing more ownership with the hiring manager. Through the recruitment process, the framework provides a useful and correctly balanced convergent divergent decision making approach for non-HR specialists to use giving ownership of the selection process through comparison of their realistic requirements as defined in the position profile and analysed in a lean recruitment Pugh matrix. As with the total design framework, where the wrong choice of concept can rarely be recouped by brilliant detailed design work, the wrong choice of candidate at the recruitment stage cannot necessarily be recouped by brilliant personnel management skills when the selected candidate is employed. The lean recruitment framework is introduced as an objective recruitment framework focusing on the value adding steps to ensuring the right candidate is employed matched to the realistic requirements of the hiring manager.

Future research should look to further develop the lean recruitment process perhaps documenting the process into an easily transferable format for generic use in the recruitment process at a wider level covering other functional requirements of a business.

REFERENCES

“The Demographic Change in Firms of Western Industry Nations - Managerial Implications for Working Systems Design”

Marion A, Weissenberger-Eibl*, Johann Schwenk

Department of Innovation and Technology Management
University Kassel
Kassel, Hessen, 34109, Germany

Abstract

The world is right now in the middle of a radical demographic transition. On the first hand the population growth is slowing down. On the other hand the age structure of world’s population is changing, with the share of younger people dropping down and that of elderly people increasing. So in most Western industry nations we have to focus a significant aging of population and therewith of human resources in firms from now to the year 2020 in a first and to the year 2050 in a second step. Cause of their psychological and physical characteristics, elderly employees underlie a different design of working systems and working conditions. In this regard it seems to be clear that firms have to challenge this changing structure of their most valuable resource to stay competitive in future economic environments.

For this attend the presented paper contributes possible managerial implications regarding the design of working systems based on the demographic change of human resources in firms. To give this objective a detailed analysis of the data of demographic change in Western industry nations, USA, Japan, Germany, and its firms are given. Related to the data analysis and the connected older growing staff in firm’s costs of this evolution are deduced. Thereto different theoretical aspects will be connected to the analysis of data. Based on these considerations basic implications for managing the demographic change are stated. Next to the central design of working systems, three other fields of related interests health, education, and culture are mentioned. Finally new illuminating aspects of future researches and developments are shown in this research context.

1. Introduction

“The world is in the midst of a major demographic transition. Not only is population growth slowing, but the age structure of the population is changing, with the share of the young falling and that of the elderly rising”, so the “International Monetary Fund”[1]. Especially for the economies in Western nations the changes are dramatic. The development in these nations will bring a significant aging of population and so to the human resources for firms from now on to the year 2020 at first and then to 2050. So Bloom and Canning from the Harvard Institute for Global Health state that connected with the prognosis of increase in global population is a complex dynamic setting of age structure, which surrounds a almost quadrupling of 60 years of age population and over by the year 2050 [2]. In this context it seems to be clear that Western industry nations and their economies as well as their profit oriented firms have to challenge this structural changing of the most valuable human resources to stay competitive in the next decades in a globalised world. Therefore demographic conditions are strongly affected by economic and social realities and institutions on the one hand, but on the other hand those realities and institutions are substantially determined contrary. This interdependency describes the underlying reflections like the two sides of one coin [3]. For the underlying research it is necessary to take a closer look by a more detailed analysis of data in combination with relevant theoretical aspects of costs. Furthermore some basic implications regarding the design of working systems, health, education, and culture are focused. The analyzed data mainly concerns the three

* Corresponding author: Tel.: (0049) 561-8043506; Fax: (0049) 561-8047029; E-mail: marion@weissenberger-eibl.de
Western industry nations USA, Japan, and Germany which are characterized by similar age structures. For some analysis the situation is focused on German data.

2. THE DEMOGRAPHIC CHANGE AND ITS COSTS IN WESTERN INDUSTRY NATIONS - THE CASE OF USA, JAPAN, AND GERMANY

The triad of USA, Japan, and Germany builds the core of analysis in the underlying research paper for two significant attributes of Western industry nations. The first attribute is that their economies are the leading head of the “Organisation for Economic Co-operation and Development” (OECD) and build therefore the most influential industrial areas. The three countries USA, Japan, and Germany are responsible for more than 50% of the contribution benefits of OECD. More detailed, the contribution in 2007 of USA is 24.975%, of Japan it sums up to 16.656%, and of Germany runs 9.265% [4]. According to this fact the called three nations realize the three highest gross domestic products in the world [5]. The second attribute concerns the fact that the USA, Japan, and Germany are facing nearly the same major demographic imbalances. In all three countries the baby boom after World War II carries a similar age structure for them. “The post-World War II baby boom and its echo (a subsequent baby boom after 25 years, as the boomers themselves had children) profoundly destabilized the age structure, for variable periods of time, in many developed countries”, so Bloom and Canning [6].

Focusing the USA we can state that the population has been relatively/high/low/balanced in the first half of 20th century. Since the 1950ies we can observe that the United States are in the middle of a dramatic demographic change which replaces the younger age structure with an older one. So a reduction of the 0-19 aged people from 33.9% in 1950 to 26.0% in 2050 accompanied with an increase of 65+ aged people from 8.1% in 1950 to 20.6% in 2050 can be expected [7]. The median age will change from 30.0 years in 1950 to 41.1 years in 2050 [8]. For Japan we can state an even more radical change in the share of elderly aged people growing. In regard to Kozu, Sato, and Inada the shift of the age-pyramid to an age-mushroom is even stronger than in the USA [9]. The data from the United Nations shows that the median age of 1950 of 22.3 years is getting up to 54.9 years in 2050 [10]. “Europe is frequently called the old continent and it truly deserves this name in a demographic sense” [11]. So the “UN World Population Prospects: The 2006 Revision” gives us a changing median age in Europe from 29.7 years in 1950 to 47.3 years in 2050. One of the strongest marked changing in the age structure of population in Europe has to be challenged by Germany. So we can state a shift of the median age of German population from 35.4 years in 1950 and over 39.9 years in 2000 up to 49.9 years in 2050 [12]. The following explanations, figures, tables and data are corresponding to the case of Germany. An adaption and transfer of theory and the analyzed data on the USA, Japan as image of Western industry nations will given in the ensuing section.

To reverse this demographic trend Höhn, Mai, and Michel state two demographic options, higher net immigration and the rising of fertility. Taking a look at the second named option the absence of potential mothers and fathers determines the shrinking German population of the next decades. In the long run a higher fertility seems to be more efficient than higher migration (Table 1 shows a much stronger impact on the age structure through fertility than through migration). In comparison within the alternative regarding higher immigration the dependency ratio increases to 85 up to 2050 rather than to 91 in the references scenario [13].
Of course this development has strong effects on firms and the economy as a whole. Taking the stated points in account two trends can be forecasted. Looking at the following table - trend 1 (Table 2) - and figure - trend 2 (Figure 1) - the two trends can be pointed out. Trend 1 is the changing population at working age, which will reduce their shares up to 2050 very intense [14]. Trend 2 of demographic change in German economy and firms intensifies the forecasted scenario. Not only that Germanys working population is shrinking but additionally that the resource workers are getting older and older. In this context the age related structure of human resources will be critically diversified towards an older population in general and so concerning the economic system to elderly employees in future [15].

Table 1: Comparison of old-age dependency ratios under alternative assumptions in Germany

<table>
<thead>
<tr>
<th>Year</th>
<th>Reference scenario</th>
<th>Alternative: Higher migration</th>
<th>Difference</th>
<th>Alternative: Higher fertility</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>45.2</td>
<td>45.2</td>
<td>0.0</td>
<td>45.2</td>
<td>0.0</td>
</tr>
<tr>
<td>2010</td>
<td>47.4</td>
<td>47.3</td>
<td>-0.1</td>
<td>47.4</td>
<td>0.0</td>
</tr>
<tr>
<td>2020</td>
<td>58.2</td>
<td>57.0</td>
<td>-1.2</td>
<td>58.2</td>
<td>0.0</td>
</tr>
<tr>
<td>2030</td>
<td>78.4</td>
<td>75.2</td>
<td>-3.2</td>
<td>76.4</td>
<td>-2.0</td>
</tr>
<tr>
<td>2040</td>
<td>83.5</td>
<td>78.9</td>
<td>-4.6</td>
<td>73.7</td>
<td>-9.8</td>
</tr>
<tr>
<td>2050</td>
<td>90.8</td>
<td>85.1</td>
<td>-5.7</td>
<td>72.9</td>
<td>-17.9</td>
</tr>
</tbody>
</table>

*Life expectancy: baseline scenario; TFR: 1.4; net migration: 100,000 per year.

bLife expectancy: baseline scenario; TFR: 1.4; net migration: 200,000 per year.

\^Life expectancy: baseline scenario; TFR: 2.1; net migration: 100,000 per year.

Table 2: Demographic Projections for the Population at Working Age in Germany
“Perhaps the most imminent problems of aging will show up on the labor market” [16]. The cause of the fact that a deep decrease of working age people will face Germany from 2015 on - as the USA, Japan and in general Western industry nations - strong influences on the gross domestic product (GDP) can be expected. The decline of people at employable age, assuming a constant employment rate, labor participation rate, as well as the number of workers would shrink equal in future, the trend of labor productivity will continue and will bring a changing demographic structure along that limits the annual growth of gross domestic product (GDP) and growth domestic product (GDP) per capita by circa 1 % up to 2020 and by circa 0.5 % between 2020-2030. So there can be a rising of indirect costs through the demographic change through the negative influenced labor market as the differences of a possible but not realized future gross domestic product (GDP) [17].

Despite to the macro level effects on economic systems we can take a more detailed look at the micro level, the firm level through at the perspective of direct and indirect costs of work-related diseases. As figure 2 clarifies there are direct and indirect costs of work-related diseases on the left side concerning a physical workload and on the right side regarding a psychic one.
As workers' direct costs we state the therapy against diseases and indirect costs as productivity breakdowns due to work-related diseases through unemployables. Physical workloads we can describe as physical strains such as heavy liftings and as psychic workloads time pressure and stress. Statistical relevant for the current analysis is that the illustrated data must not be aggregated, because they are not independent from each other. We can suppose rising costs of work-related diseases through aging staff within Germany up to 2050, when we adopt the often proven hypothesis, that the older human resources grow the bigger the chances of getting limited by most of the relevant diseases, as the starting point of the underlying working context.

Taking a detailed look at the data, figure 3 shows that one third of all work related diseases emerge through ailments in the skeleton and muscular system. Furthermore looking closer at the data we can see a strong position of the back of the personnel. Beyond these respiratory organs and psychiatric diseases are shown to be relatively high in comparison to the rest of the analyzed data. For the reviewed diseases we monitor an increase accompanied by an older growing workforce for firms in Western industry nations. Especially in the direct areas of production an over proportional enlargement of diseases of skeleton and muscular systems as well as back disorders can be validated in regard to elderly workers compared to younger ones. For that reason of relative similar correspondence in age structure and the changing of the population we can project similar evolutions. Specially the mentioned cases above of the US, Japan and Germany are examples addressing other Western industry nations [18].

![Figure 3: Diagnosis Specific Direct and Idirect Costs of Work-Related Disease in Germany in 1998 (in Billion €)](image)

"Demographic change as it is nowadays observed in many developed countries is largely unprecedented. There are thus certain limitations to systematically analyzing its economic and fiscal consequences." Making the next step for the underlying research paper the analysis of macro and micro level costs of Western economy systems and its firms a second step in that direction will be made through some basic implications for managing the demographic change in organisation of Western industry nations. In the following section the critical points of designing working systems, health, education, and culture are the central challenging topics. Although the analyzed data mainly concerns the actual situation in Germany, some interesting implications for the USA, Japan and Western industry nations in general can be made.
3. BASIC IMPLICATIONS FOR THE MANAGEMENT OF DEMOGRAPHIC CHANGE

In regard to the discussion of the economic conditions of demographic changes in firms of Western industry nations there are two ways of how to continue. Either the addressed countries will follow the approached path of youth oriented personnel and corporate strategy or they challenge the systematically development of elderly but anyhow with competitive units. It seems to be obvious that the first way of going about it can be stated as deadlock. Following the first option, as shown within the analyzed data and cost analysis above, will bring next to the shortage of skilled workers a negative age-related impact on the productivity of firms and in all Western industry nations. So the only way out of this deadlock seems to be the second option. The critical question is how to develop systematically older growing but competitive firms further on [19]? 

The first two starting points can be directly answered through the analysis above. To the fact that the biggest amount of work related diseases can be detected in the fields of skeleton and muscular system as well as in the field of back disorders, the possible implications for management illuminates, 1) the ergonomic, age-oriented design of working systems, and 2) the active promotion of the staffs health.

1) In the ergonomic, age-oriented design of working systems we have to differ between the direct and indirect divisions of firms. The direct area refers to the more relevant and bound costs, the production area to firms. Possible objectives to design the working systems more age-oriented are to reduce critical sizes like heavy lifting, frequent bending and constant work footing as well as giving the relevant workers lifting pins. But also the situation in the indirect area, the administrative departments, can be improved through ergonomic changes. To first thing thinking of is to give the employees an adequate work place like appropriate office chairs, desktops and monitors. All in all the huge field of ergonomic designing of firms in the direct as well as the indirect divisions is essential. Everything considered as general objective is to reduce working places, which are posing a risk to the health of workers and employees for an older growing but competitive firm.

2) The second statement pointed out above is to generate a proactive health care in industrial firms. Therefore an effective health and safety protection at work seems to be crucial in this nexus. Regarding this the focus lays on the prevention of health management and disease. On the one hand we can state the valuation of working conditions here as a basis for a health-oriented and age-oriented design of working places / processes as relevant. On the other hand it is important to inform the staff about health risks. Add itional trainings for health-oriented behaviour is significant e.g. how to lift and to carry things. Constantly provided health checks for the staff, a regularly accomplished supervising through a company’s medical office and the active support of sport and wellness activities round of the possible options to the demographic change and health care. Managers should bear in mind that the importances of health care on work and in and with the ergonomic, age-oriented design of working systems.

Despite to the direct influences of managing the demographic change and reducing costs of older grown personnel, two more points lay open in literature to attack, 3) an age-related education in terms of long life learning, as well as 4) a corporate culture, which sees older employees as a benefit and not as a thread for staying competitive.

3) Next to the already stated points strong in vestments in the age-related education of firms’ employees is a profitable point. Tw o different but complementing ideas of e ducating concepts have to be discussed a s the long-life learning is not an empty phrase. Ad equate fo rms of f urther e ducation and training an d ex panding the current knowledge pool organized for older growing people help firms to stay competitive. In addition special forms of trainings, which support the work experience knowledge of older employees seems to be important. Here we suggest some intra-firm tools like “lessons learned”, “communities of practice (COP)”, an d knowledge-based exchange platforms as meaningful. To maintain the experience knowledge of older people is an important task for the firms s it i s extraordinary important that competitive success should be systematically managed for a sustainable future. The second complementing point is to adapt further education and training to the challenges and special conditions of older people. The rising complexity of the environment in combination with the constantly increasing information exchange has to be thematized in connectivity with age-oriented forms of learning.

4) According to our analyzes as well as to theoretical works we called the continuation of the approached path of youth oriented personnel and corporate strategy in the current situation of demographic changes as deadlock. Nevertheless most companies and industries do not act in this direction. But firms will be forced to rethink their behaviour in the next decades. To manage demographic change in total it i s absolutely necessary to generate a
pattern of basic assumptions [20], a mental programming of people [21] that sees elderly staff not as a threat but as a chance to stay competitive in future economic systems of Western industry nations. For cultural development in this direction socialization through company’s management, which would give an positive example, is the crucial starting point. To follow age-oriented personnel, corporate strategy and to communicate and live it in firms is the only way to start with this important point.

For the four called possible starting points the actual and future demographic situation, the above conducted analysis of data and costs, considerations in the scientific theory [22] as well as institutions as the “International Monetary Fund” (IMF) and the “Organisation for Economic Co-Operation and Development” (OECD) gives evidence, so that they can be seen as fruitful influencing variables to manage the demographic change in firms of Western industry nations.

4. SUMMARY

The main objective of the underlying paper was, based on the analyzes of the data regarding the demographic development and connected costs, to give firms possible starting points for managing the demographic change of Western industry organisations. Looking at data of the USA, Japan, and especially Germany the analyses gave us an idea how radical the demographic change and therefore common challenges for Western industry firms will be. Deducing costs out of this basis shows the importance for firms in Western industry nations to manage this evolution systematically for staying successful and competitive in future. Regarding these costs which mainly result out of work-related diseases e.g. of the skeleton, muscular system and back disorders some basic implications for firm’s management for demographic change can be given. 1) The ergonomic, age-oriented designing of working systems, 2) the active promotion of health, 3) age-related education and 4) a corporate culture, which sees older employees as chance and not as a threat were argued as possible answers to the question how to develop systematically older growing but competitive firms.

With this paper we made a critical step how to manage the actual and future noticed demographic change in Western industry firms, to get a human resource base which belongs to the different age structure to gain sustainable competitive advantage. Although the data of Germany is similar for the situations in the US and Japan, the results cannot be simply transferred in general, but needs to be strengthened in this context. Furthermore a more detailed look at the analyzed data can be seen as future research field to get new and more proven insights in general and especially in their connectivity to one of each other. Last but not least more detailed work for the four suggested determinants of managing the demographic change working systems - health care, education, and corporate culture - has to be done. On the one hand they have to be developed more sophisticated and on the other hand their reciprocal dependences must be analyzed with a more focused look.

REFERENCES


Analysis of the Methods Time Measurement (MTM) Methodology through its Application in Manufacturing Companies

Denis L. M. de Almeida¹, João C. E. Ferreira¹

¹Departamento de Engenharia Mecânica
Universidade Federal de Santa Catarina, GRIMA/GRUCON, Caixa Postal 476
Florianópolis, SC, 88040-900, Brazil

ABSTRACT

With the continuous growth of competitiveness among companies, the increase in productivity has become an important requirement for many of those companies. However, it is important to think about alternatives that allow the increase in productivity without overloading the resources, avoiding equipment errors and health problems of the operators. In order to increase productivity, this work proposes the use of the Methods Time Measurement (MTM) methodology. This work analyzes the application of this methodology in manufacturing companies considering the current developments of time tables, which are not widespread in many companies. However, in this work the MTM methodology will not be applied in an isolated way, but in combination with other techniques. Details are given with regard to the methods and results obtained through the application of MTM in two companies of the automotive sector, and a company that manufactures household appliances. Finally, it is concluded that the MTM methodology is a useful tool for the planning and organization of operators’ working processes, and that combined with other applied methods it can generate a significant increase in productivity.

1. INTRODUCTION

We live in a world with scarce resources, and society is less and less accepting to pay the high price of waste. In order to be successful, companies need more and more to eliminate waste so that products with greater quality, lower costs, and shorter production times are provided to customers. Besides the high competitiveness among companies and the ever-increasing level of customers’ demand, the increase in the conscience of the workers about their contribution to the final product lead to a scenario where high productivity and responsiveness levels are mandatory. Thus, it is necessary the development and application of techniques that allow the best utilization of the available resources.

In order to increase productivity, this work proposes the use of the Methods Time Measurement (MTM) methodology in different companies.

MTM [1] defines the MTM methodology as an instrument to describe, structure, configure and plan work systems through defined process modules, seeking to be an efficient pattern of production systems. It can be used anywhere where it is necessary to plan, organize and accomplish a human task seeking its effective execution. Basically, "the method determines its time". When used for planning issues, the MTM methodology would justify the premise presented by MTM [1], which is to "avoid costs instead of reducing them", that is, a correctly planned process can be executed since the beginning without the extra costs associated with the inefficiencies of the process.

However, in spite of being a very useful tool, the MTM methodology, in its current stage of development, is still little known in countries like Brazil, which generates inadequate interpretations regarding its application. According to Priemer [2], manufacturing engineering and workers in Brazil, when they refer to the MTM methodology, make the following comments: "it is just a Taylorist method", "we don't know about it", "it is not applied in our production", "it is applied at the head office (abroad), but not here", "it is very detailed and it needs an excessive effort", "maybe it can bring us advantages."

* Corresponding author: Tel.: +(55)(48) 3721-9387 ext 212; Fax: +(55)(48) 3721-7615; E-mail: jcarlos@emc.ufsc.br
2. METHODS TIME MEASUREMENT - MTM

The MTM methodology is a method to structure sequences of movements into basic movements [3]. The value of a standard time is assigned to each basic movement, which is (pre) determined as a function of the factors that influence its composition. The MTM methodology can be applied to: configuration of the work methods and products; determination of times; to describe the method as a form of documentation for training [3].

The MTM methodology is based on the following five basic movements: to reach, to grasp, to move, to position and to release (see figure 1) [3]. These movements compose 80% to 85% of the procedures wholly influenced by human beings. Besides these basic movements, the following actions are used to describe the movements: to apply pressure, to separate, to twist, body movements and visual functions. The body movements are divided into: foot movements, leg movements, lateral step, body spin, walk, bend, leave the bent position, lower, leave the lowered position, kneel on one knee, leave the kneeling position on one knee, kneel on two knees, leave the kneeling position on two knees, sit down, leave the seating position.

The MTM methodology is in continuous development. Some of the world developers of this methodology, always based on the basic method, are as follows:

• GPD (MTM-General Purpose Data - 1963) – American MTM Association;
• MTM-2 (1966) – Swedish MTM Association, being used in Scandinavia, England and France;
• MTM-SD (Standard-Daten-Basiswerte – Basic Values “Standard Data” MTM);
• MTM-UAS (Universelles Analysiersystem – Universal Analysis System);
• MEK (MTM für die Einzel- und Kleinserienfertigung – MTM for individual production and for small series).

3. METHODOLOGY

According to Sugai [4], the process of applying MTM should be considered as a process of continuous improvement, and it points out six stages for its application in a work station and four stages for its application in a company as a whole, which can be summarized in figures 2 and 3, respectively.
In this work case studies of the application of MTM in three companies were analyzed, being two car companies and a company that manufactures home appliances. A German plant that manufactures cars was evaluated, whereas the other two companies were evaluated in Brazil.

For the accomplishment of the MTM analyses, the following data sources were used: (1) direct observation of the work station by MTM analyst; (2) analysis through films; (3) documents (such as process plans and technical data about the products); (4) simulation of the movements by the MTM analyst; (5) experience of the operators and process planners. The choice of which sources would be used depends on the availability of these data and time to accomplish the analyses.

Once the data is collected, they were evaluated through the TiCon software [5], which was also used to balance the activities. However, this work was not limited to accomplish time analyses, but it sought to combine the MTM methodology with other methods to allow the productivity increase in the evaluated areas.

The work to increase productivity consisted in: (a) to determine the correct times for each activity; (b) to study the work stations seeking for better solutions and methods; (c) to redistribute the activities in order to level the line; (d) to propose improvements; (e) to priorize and implement improvements; (f) to train operators; (g) to validate the proposals. Besides the stages above, audits correspond to an additional stage that can be applied in any of these stages to guarantee the application of the proposed methods.

It was verified that the greatest potential of the MTM methodology is in the possibility of combination with other methods, since through the detailing of the activities failures and waste that would pass unnoticed by other techniques would be identified by MTM. In this study, some techniques were verified in practice that could be combined with the MTM methodology, which helped in the search for better methods and improvement proposals, and these techniques were:

1. **To standardize:** Some of the main patterns verified in the evaluated car companies are: work in one takt; fixed cycle times; reference point in the products; movement in triangle; materials made available in the station area; avoid blocking the movement; alarms of quality problems and alarms of line stoppage; use of limited equipment
within the station; target of percentage of added-value activities; avoid processes that require more than one operator; transportation devices to guarantee ergonomic stations; use of "andons."

2. **Layout:** Besides allowing the improvement of the transport, the improvement of the layout is directly linked to the ergonomic improvement of the work stations.

3. **Use the data supplied by the MTM methodology:** from the analysis of the information supplied by the MTM methodology, questions can be made in order to verify and implement improvements.

4. **Study the processing/waiting times:** the time that an operator waits for a machine to accomplish its process is waste, and as such it should be eliminated.

5. **Prevent a local improvement from worsening the global situation:** because the MTM method focuses on the local activities.

6. **Use the experience of the operators:** the fact that the operators work directly at the place to be analyzed and they experience the problems daily makes them a source of valuable knowledge for the company.

7. **To outline the process:** an outline of the process can be used as a summary of the relevant information to allow a fast and simple visualization.

8. **To motivate the search for group solutions:** The work in groups is extremely useful to find solutions for improving the process.

9. **Support employees:** the multifunctional operators are very important for the production lines, since they significantly help to level the production lines by the elimination of disturbances.

10. **Another means to optimize a process:** The proposals listed above correspond to only some of the possibilities of improving a process, but to limit these techniques into a list, no matter how complete it is, it can be considered inadequate, since the human creativity is limitless, as well as the possibilities to improve a process. Therefore, the techniques pointed out here should be used as a guide and not as a general rule to improve the process, since many other possibilities may be implemented successfully.

In the three analyzed companies, several situations could be evaluated through MTM, such as: alterations in the production volumes; activities that combined line-based and cell-based work stations; quality control, finishing, and rework activities; activities with multiple combinations of variants, logistics, etc. The MTM methodology led to good results in the evaluation of all these cases, but two cases that recurred in all the companies considered, which will be highlighted here, were the following: the comparison between MTM and times obtained though chronometer measurement analysis, and the analysis of simple/repetitive activities.

4. **CASE 1: COMPARISON OF MTM WITH CHRONOMETER MEASUREMENT ANALYSIS**

The comparison between the times obtained by the MTM method and those obtained with the chronometer measurement analysis took place several times during the MTM applications considered in this work. The following example presents the result of a comparison between a MTM analysis and the result of a simple chronometer measurement.

A versatile operator acts at three work stations "A", "B", and "C", each one with cycle times equal to 90 seconds, calculated based on the production demand. The supervisor responsible for the area, not satisfied with the results of the MTM analyses, decided to time the operator to verify the times. By simply timing the operations, but not performing a complete chronometer measurement analysis, the supervisor declares that the MTM times were not in agreement with the real situation, and he mainly criticizes station C, where the time of MTM analysis was almost 10 seconds lower than the timed one. Table 1 shows the results described above.

<table>
<thead>
<tr>
<th>Work station</th>
<th>MTM Analysis</th>
<th>Chronometer Measurement Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>95.3 s</td>
<td>90.5 s</td>
</tr>
<tr>
<td>B</td>
<td>89.4 s</td>
<td>91.0 s</td>
</tr>
<tr>
<td>C</td>
<td>71.8 s</td>
<td>81.5 s</td>
</tr>
<tr>
<td><strong>Total Time</strong></td>
<td><strong>256.5</strong></td>
<td><strong>263</strong></td>
</tr>
</tbody>
</table>
When analyzing the data, the supervisor also questioned the fact that the MTM time for station A was longer than the MTM time for station B, while the result for the chronometer analysis was the opposite.

The MTM analyst then asks the supervisor regarding his/her opinion about the rhythm of work of the operator in the work stations. The supervisor replies: “In station C, the operator worked calmly, in station B in a good rhythm, but he was accelerated in station A”. The analyst then suggested the analysis of the data through the Operator Balance Chart of the work stations.

By observing table 1, it can be noticed that in the chronometer analysis measurements the times tend to the of 90 seconds line that refers to the cycle time. The reason of this phenomenon is related to the available time for the operator to accomplish the task. The operator has 90 seconds on average to perform the activities in each station, i.e. he/she will have to make a greater effort and to accelerate the rhythm to meet the required production in station A, but he/she reduces the speed and works in a more relaxed manner, in order to take advantage of the time in work station C. In station B, the measurements with chronometer and the MTM analyses reached similar values, and the main reason is the fact that the cycle time is approximately equal to the time of the activities, i.e. a correct MTM analysis will result in values very close to 90 seconds, and the operator will have to maintain a good rhythm to meet the needs of this work station. It can be inferred that the MTM analyses emphasize the different workloads at the station, since they do not suffer influence, for instance, of the pressure on the operator to accomplish his/her task in the correct time.

It should be emphasized that, if the chronometer analysis were performed adequately, repeating the measurements in the appropriate amount and applying the rhythm evaluation, the times would approach the times obtained by the MTM method, which can be verified in cases where the chronometer analysis was applied correctly.

5. CASE 2: ANALYSIS OF THE SIMPLE/REPETITIVE ACTIVITIES

This topic is based on the need to accomplish correct analyses in the shortest possible time. Britzke et al. [6] perform a comparison between the analyses accomplished with the GMA, GMA 2 and GMA 3 methodologies, which are appropriate modules for mass production, series production, and small series production respectively. This comparison can be extrapolated as reference for the modules MTM-1, SD, UAS, and MEK (pointed out in section 2), considering the following combinations: (a) GMA and MTM-1 for mass production; (b) GMA 2 and SD/UAS for series production; (c) GMA 3 and UAS/MEK for producing in small series.

Britzke et al. [6] point out that, for the GMA, GMA 2 and GMA 3 modules, it takes from 200 to 300 minutes to analyze 1 minute in the case of mass production, from 100 to 150 minutes for series production, and from 20 to 30 minutes for the producing in small series. In other words, the time necessary to perform an analysis for production in series should be approximately 50% of the time required for the analyses of mass production, and the time for small series should be equal to 10% of the time for mass production.

Those values should be considered only as reference, because it can be verified, when accomplishing the analyses, that there are several factors that contribute to reducing (or to prolonging) the analysis time, such as the following: (a) experience of the analyst with the methodology; (b) the knowledge of the analyst about the time tables to be used, as well as the time tables that served as reference; (c) the knowledge of the analyst about the stations to be analyzed; (d) the knowledge/experience of the analyst about similar work stations; (e) access to information (p.ex., film, direct observation, access to the operators, historical data, documentation, procedures with descriptions of the station’s activities); (f) level of the used method; (g) support from the people involved, and (h) the analyst’s personal factors (motivation, diligence, etc.).

Besides these factors, an aspect that influences significantly the speed of the analyses is the degree of complexity of the work station. For instance, an inexperienced analyst would probably have great difficulty in analyzing activities of finishing or those having multiple variants.

On the other hand, there are analyses that become simple due to the repetition of movements, which speeds up the execution of the analysis. A typical example of this type of analysis is the analyses of the work on press machines. Despite this type of work occurs in several different applications, it can be affirmed that in many companies such movements can correspond to 80% of the production movements (disregarding the maintenance activities, setup, etc.) when considering production sectors/teams dedicated to these activities.

The typical movements in this process are: (a) to grasp parts; (b) return movement from grasping parts; (c) to position the parts in the press; (d) to depress the bi-manual command (due to the existing risks and demands of the standards, the bi-manual command ends up being one of the most common cases of actuation); (e) process execution; (f) to grasp parts; (g) to deposit parts, and (h) return movement from depositing parts.

Table 2 presents an analysis of activities of the work in presses, based on the cases verified in practice.
Table 2: Example of the analysis of the activity in Press, analysed with MTM UAS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grasp sheet with 400x400mm on the side, and position in press</td>
<td>AJ3</td>
<td>2.7</td>
<td>Grasp bulky part and position it without precision</td>
<td>1</td>
<td>1</td>
<td>2.7</td>
</tr>
<tr>
<td>Complete movement to grasp parts in a 2m area</td>
<td>KA</td>
<td>0.9</td>
<td>Move / meter</td>
<td>2</td>
<td>2</td>
<td>3.6</td>
</tr>
<tr>
<td>Depress bi-manual command</td>
<td>BA2</td>
<td>0.9</td>
<td>Simple drive, distance between 20 and 50 cm</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Execute the process – Press in operation</td>
<td>PT</td>
<td>1.0</td>
<td>Processing time in seconds</td>
<td>1</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Grasp pressed sheet and put it in the pallet on the floor</td>
<td>AJ3</td>
<td>2.7</td>
<td>Grasp bulky part and position it without precision</td>
<td>1</td>
<td>1</td>
<td>2.7</td>
</tr>
<tr>
<td>Bend and return to upright position to put pressed sheet in the pallet</td>
<td>KB</td>
<td>2.2</td>
<td>Bend and return to upright position</td>
<td>1</td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>14.1</strong></td>
</tr>
</tbody>
</table>

6. MTM COMBINED WITH OTHER METHODOLOGIES

There are many methodologies that can be applied in parallel with the MTM methodology, and some of them were applied during the studies of this work, such as the kaizen workshops and the 5S philosophy. The main conclusion regarding this topic is that the MTM methodology is not competitive with other managerial tools, but complementary. The chronometer analysis, which many people consider as being competitive with the MTM methodology, is actually a good example of method that can be applied together with MTM.

In these cases, the chronometer analysis can be used as a counterevidence of the times evaluated with MTM and vice-versa. Another great possibility of combining these methodologies is the application of MTM to activities influenced by man, and the chronometer analysis for processes that depend on machine times. This logic can also be extended to other time measurement methods, such as the sampling and/or estimates of times that can be used, for instance, to evaluate the times that the operator needs to go to the bathroom, to drink water, etc.

Regarding the planning situations, the MTM methodology was very advantageous, and its use is strongly recommended as main method, but for series activities it is recommended the combination of MTM with other methods for time evaluation.

A frequent criticism to the MTM methodology by the people involved with analysis of times in companies is the time necessary to accomplish the analyses, but this point was proved wrong during the execution of this work. To avoid the long execution times, it is important to structure the analyses and the database of the production times of the company, which is usually accomplished in the chronometer analysis. This procedure enables the reuse of the data in future situations. With that structuring, it was verified that the MTM analysis can, depending on the situation, be faster than the chronometer analysis.

Considering only the MTM analyses in each day worked, the following was analyzed:

- Little more than 30 seconds of activities of the operators a day in the first weeks of performance, which is considered a low analysis speed, mainly due to the unawareness of the tables used in the considered company;
- Approximately 1 minute of activities of the operators a day, including the practical leveling (practical application of the results in the production line) in the following weeks;
- 6 minutes of activities of the operators a day after some months of applying the MTM methodology. This duration can be considered appropriate for analyses by trained MTM analysts, for the cases in which the company’s structure of times is ready.
• 12 minutes of activities of the operators per day in favorable cases after some months of applying the methodology.

For instance, comparing MTM with the chronometer analysis, which needs to measure 30 repetitions of the activities, in order to evaluate 12 minutes it would be necessary 360 minutes of direct measurement, i.e. 6 working hours, which implies that analyses such as rhythm evaluation and dismissal of measurements with errors, would need to be accomplished in little more than 2 hours to have a rhythm equivalent to the MTM methodology, not considering the data structuring for a subsequent reuse.

An important observation to be made is that these times were analyzed for serial activities, whose cycle time was equal to approximately 1 minute, and the time of method was equivalent to the SD / UAS tables. Different applications require different times for analysis.

Another observation is that the times presented here do not consider the bureaucratic processes of the companies, such as the need to report the analyses to the high management or collecting several signatures of specialists. These times were frequently superior to the times of analysis, and they should be taken into account when planning the application of MTM.

Similarly to the chronometer analysis, Lean Manufacturing can also be combined with the MTM methodology, because both have as objective the reduction of costs through the elimination of waste, but while the Lean Manufacturing focus the processes of the companies as a whole, the MTM methodology focuses on the manufacturing processes, with emphasis in the human activities.

An important issue considered by Lean Manufacturing that was applied in practice through the MTM methodology is the classification of the activities into value-added activities, non-value added activities (but are necessary), and waste. Another methodology whose combination with MTM can lead to significant gains in productivity is the rapid exchange of tools, as pointed out by Cakmakci and Karasu [7].

Besides the methods pointed out above, the MTM methodology was combined with other methodologies during the execution of this work, all with good results. Some of these methodologies were: 5S, Ergonomics, Continuous Improvement, PDCA Cycle, Kaizen Workshops, Benchmarking, SPC – Statistical Process Control, etc.

7. CONCLUSIONS

Regarding the developed work, the first observation is that the MTM codes and times are universal and guaranteed by IMD (MTM International Directorate). The practical observation proved that the times that are valid in Brazil are also valid in Germany, and this can be extrapolated to any part of the world, which can be proven through international contacts.

The MTM code helps significantly the understanding of the processes by somebody who does not dominate the language in which the descriptions are made. Somebody who knows the used code and can visualize the described process is capable of understanding the analyses even without understanding a word of the descriptions of the activities.

It could be verified that in a general there is a greater acceptance of the MTM methodology in Germany compared to Brazil. In Brazil, the MTM methodology is ignored in general by great part of the companies. One of the facts that can be used as proof of the greater acceptance of the MTM methodology is the fact that even in the companies that use MTM in Brazil, few people know the methodology, while in Germany this methodology is in general more widely known and applied.

When the MTM methodology is analyzed in an isolated way, it can be considered as a table of times, and as such it does not increase productivity. However, the need for detailing the tasks for its analysis process enables the identification of small wastes, which pass unnoticed through other improvement processes, and consequently MTM provides an appropriate base so that improvements may be implemented. The need for detailing the tasks contributes to reduce the effort required to accomplish the analyses, which was minimized with the most recent developments of the MTM methodology, where time tables were developed for different levels of methods.

After accomplishing this work, we conclude that the MTM methodology is a useful tool for planning and organizing work methods of the operators, and when it is combined with other methods applied by the companies it can lead to a significant productivity increase.

Also, the following additional conclusions can be pointed out:
• The operators should not work during more hours, but in an efficient way, in other words, the MTM methodology does not intend to reduce the time that the operator has for accomplishing a certain activity, but to eliminate the unnecessary movements that waste time;

• The main factors associated with the success of the implementation of the MTM methodology verified in this work were: company policy, support from company management, structuring of the data, and training of the operators;

REFERENCES

A Neural Network Approach for Measuring the Impact of Knowledge Management on Organizational Performance of Turkish SMEs

Ali Turkyilmaz\textsuperscript{1}, Erkan Bayraktar\textsuperscript{2,*}, Halil Zaim\textsuperscript{3} and Selim Zaim\textsuperscript{3}

\textsuperscript{1}Department of Industrial Engineering, Fatih University, B.Cekmece, Istanbul, 34500, Turkey
\textsuperscript{2}Department of Industrial Engineering, Bahcesehir University, Besiktas, Istanbul, 34349, Turkey
\textsuperscript{3}Department of Business, Fatih University, B.Cekmece, Istanbul, 34500, Turkey

\textbf{ABSTRACT}

Knowledge management (KM) has recently emerged as a discrete area in the study of organizations and frequently cited as an antecedent of firm performance. Managing the knowledge resources effectively and efficiently seems to be the only way to gain continuous and sustainable competitive advantage for organizations in the long-run. However, the concept of KM is still evolving and therefore, it is not easy to develop a theoretical framework that has been widely accepted. This paper analyzes KM practices based on eight dimensions including organization structure, technological infrastructure, organizational culture, intellectual property, knowledge generation, knowledge codification and storage, knowledge transfer and sharing, and knowledge utilization. A customized neural network model was developed to test the theoretical model based on a sample of 277 small and medium sized enterprises (SMEs) operating in textile industry of Turkey. Empirical analysis of the model showed that there was a strong and positive relationship between the level of KM practices and organizational performance. Further, managerial implications to improve the performance of SMEs were proposed in terms of KM practices.

1. INTRODUCTION

As one of the contemporary management tools, knowledge management (KM) has been increasingly used by most of the large-sized organizations and multinational companies to gain sustainable competitive advantage in the long run. Despite its growing implementation, the concept of KM is still evolving and, yet there is no unifying or overarching theoretical framework that has been widely accepted.

While KM has been frequently cited as an antecedent of organizational performance, there is a paucity of empirical research regarding the impact of KM practices on organizational performance. This lack of interest is even more pronounced in the context of small and medium-sized enterprises (SMEs). While implementation of KM practices in large size firms provides immense business opportunities in terms of achieving cost efficiency and gaining competitive advantage, there is less evidence of small- and medium-sized enterprises (SMEs) implementing KM practices to capture similar benefits. The question then remains open how well KM practices fit with the SMEs, which form the largest group of manufacturing firms in both developed and emerging market economies from the viewpoint of generating employment and economic growth (Demirbag et al., 2006). They account for more than half of the employment and value added in most countries (UNCTAD, 1993). Similar trend is also observed in Turkey where SMEs constitute 99.5 per cent of all business establishments and employ 61.1 per cent of the workforce (Yilmaz, 2004).

In view of the fact that the success of SMEs has a direct impact on the national economy, this study provides essentially two main contributions to SME research. First, based on a sample of SMEs operating in two sub-sectors of textile industry within the greater metropolitan area of Istanbul in Turkey, this study aims to examine the impact of KM practices on the organizational performance of SMEs. Second, the neural networks approach which has been gaining recently a growing interest in business research is also employed to identify the most important KM practices on organizational performance of SMEs.

* Corresponding author: Tel.: (90212) 381-0566; Fax: (90212) 381-0550; E-mail: erkanb@bahcesehir.edu.tr
The remainder of the paper is organized as follows. The next section briefly reviews the relevant literature on KM practices. Research methodology is presented in Section 3. Data analysis and results are provided in Section 4. Conclusions are in the final section.

2. LITERATURE REVIEW

The field of KM has recently emerged as a new area of interest for both academic and business circles. The literature reveals an increasing number of studies covering many different fields of KM (McAdam and Mcreedy, 1999). Along with this growing interest, there is a vast number of definitions with more or less common in characteristics (Lytras et al., 2002) emphasizing several different aspects of KM. The existing studies in the field of KM have largely focused on three major streams (Geisler, 2007): the nature; the processes of knowledge generation, sharing and distribution; and infrastructure for technology and effective management of knowledge. Zaim et al. (2007) classify the infrastructure further into four areas: technology, organizational culture, organizational structure and intellectual capital. Similarly they also identify four areas of processes for KM: knowledge generation and development; knowledge codification and storage; knowledge transfer and sharing; and knowledge utilization. In the forthcoming section, we will develop the concept of KM in line with the categorization purported by Zaim et al. (2007) and will subsume both KM processes and related infrastructure under the general heading of KM practices.

2.1. KM PRACTICES

It has been argued that the effectiveness of KM depends on how the generation of new knowledge is organized and how existing knowledge is transferred throughout the organization. Recent studies have expressed considerable interest in knowledge sharing practices (Hicks et al., 2007). The benefits of knowledge transfer and sharing have also been discussed widely among the scholars and practitioners (Sveiby and Simons, 2002). Therefore, one of the most important objectives of KM is to bring together intellectual resources and make them available across organizational boundaries. It has been suggested that organizations often waste their resources and lose a significant amount of money for repeating the same mistakes, duplicating projects and being unaware of each others’ knowledge due to the lack of knowledge transfer and sharing throughout the organization (Robertson, 2002).

Knowledge transfer is not a unidirectional movement. Effective knowledge transfer is more than the movement of knowledge from one location to another. Organizations can get significant learning experience through knowledge transfer between units and people. It tends to improve competency of both sides that transfer and share knowledge. It is because knowledge does not leave the owner when it has been transferred. As a result, the value of knowledge grows each time a transfer takes place and the key to value creation lies in how effective knowledge has been transferred throughout the organization.

The role and importance of information and communication technologies in knowledge transfer have been emphasized by many scholars. Clearly, technological advances bring a vast number of new opportunities to transfer and share knowledge and expertise throughout the organization within departments, plants, countries and across national borders. These technologies have a strategic role in knowledge sharing specifically for the geographically dispersed global organizations (Bender and Fish, 2000). The effective use of technologies creates new ways of knowledge transfer and hold promising solutions both in transfer of explicit knowledge and tacit knowledge – in terms of experience and expertise- (Jacop and Abrahampur, 2001). In this respect, it is often mentioned that technological infrastructure has a strategic importance in knowledge transfer not only within the organization but also among different organizations (Zhao and Xie, 2002).

As a matter of fact, all healthy organizations generate knowledge. While they are interacting with their environment, they absorb information, combine it with their experiences, values and internal rules, turn it into knowledge, and take action based on it. Knowledge generation can be performed in many ways. The three of the main modes amongst others are knowledge acquisition, knowledge generation within the firm and collaborative knowledge generation. However, knowledge generation process is a set of activities for the conscious and intentional generation of knowledge under specific actions and initiatives firms undertakes to increase their stock of corporate knowledge (Davenport and Prusak, 1998).

Knowledge generation process does not necessitate new knowledge generation. In many circumstances, organizations may prefer to acquire knowledge from other sources and adopt it for their own use (Bhatt, 2000). Knowledge acquisition can be used for knowledge creation, and if it is novel and useful for the organization, also be considered as a part of knowledge generation. Organizations convert information they collect from internal and external sources into knowledge through their organizational learning process by combining it with their prior
knowledge, experiences, values and organizational procedures (Hong, 1999). Then, the knowledge becomes a part of their organizational knowledge base. This obviously explains why the knowledge acquired through these organizational processes is new and unique for that organization (Krogh and Roos, 1996).

Knowledge is meaningful when it is codified, classified, put in a useful format and stored. Only then, it can be used by the right person, at the right time and in the right way. Knowledge codification and storage is important not only for an effective use of knowledge but also for reusability of knowledge in case it is needed so that the knowledge in question can be internalized to the organization rather than the knower (Nemati, 2002). Therefore, considering the organization’s overall objectives and priorities, many studies have been concentrating on the classification and the codification of knowledge based on its types and purposes (Lueg, 2001), and on the storage of knowledge to let the employees be able to access knowledge any time both at present and in the future. The codification of knowledge also enables to stock the knowledge resources and to assess the potential of the organization. The most challenging feature of knowledge codification is to gather it without losing distinctive properties which makes it valuable (Davenport and Prusak, 1998).

Despite its importance, codifying and classifying knowledge is not that simple since it relies heavily on what people know. Thus, organizational knowledge is hard to capture, clarify and express perfectly fine considering the fact that it is dispersed and scattered throughout the organization. It is found in different locations, in peoples’ minds, in various organizational processes, in corporate culture embedded into different artifacts and procedures and stored into different mediums such as print, disks and optical media (Bhatt, 2001).

There is a distinction between tacit and explicit knowledge in the storage of knowledge. Explicit knowledge can be easily collected, documented, stored and retrieved quite independently of any single individual through technological means and systems. On the other hand, tacit knowledge resides in the minds of the employees and seizes a great deal of an organization’s knowledge resources (Drott, 2001). If the organization’s knowledge resources have been described as an iceberg, the explicit knowledge is the visible part of the iceberg above the surface, whereas the tacit knowledge includes the invisible part of the iceberg beneath the surface (Herrgard, 2000). The codification of tacit knowledge unlike explicit one is the most cumbersome activity in the overall process because of its subjective and situational nature, and it is intimately tied to the knower’s experience.

One of the most important and challenging aspects of KM is to enhance the development of a collaborative, trustworthy, emphatic and helpful organizational culture. The executives and scholars agree on the importance of a knowledge-friendly culture for the success of KM (Hauschild et al., 2002; Skyrme, 1999). It is because knowledge is a context-dependent social concept (Lang, 2001) and a large part of organizational knowledge is embodied in social processes, institutional practices, traditions and values (Fayard, 2003; Boisot, 1998). Therefore, no matter how powerful the tools and functions of KM are, it is of no use without willing participants and a supportive social and cultural environment (Koulopoulos and Frappaolo, 1999). While the cultural resistance is generally cited as one of the most important barriers to an effective implementation of KM (Sveiby and Simons, 2002), it is still contemplated as the neglected or underestimated side of KM practices. Therefore, it is strictly recommended for organizations to place a special emphasis on the social and cultural issues for the successful implementation of KM practices (Bhatt, 2001).

The appropriate organizational structure and guidelines as well as technical and non-technical expedients of which the organization has disposal constitute another building blocks of KM infrastructure (Beijerse, 1999). Nonetheless, there is no single appropriate organizational structure for KM. Some scholars suggest a radical re-design for KM (Malhotra, 2000), while others think that it is not necessary. However, instead of highly centralized, control-based and rigid hierarchies, more flexible, decentralized and trust-based organizational structures with empowered workers are highly recommended in the KM literature (Maier and Hadrich, 2006; Malhotra, 2005).

One of the most important objectives of KM is to create value from organization’s knowledge resources so that the knowledge held by the company can be transformed to fields of application and action (Ordaz et al., 2004). This implies the effective and efficient knowledge utilization for the organization’s competitive edge. For that reason, it has been argued that the success of KM practices mostly depends on how efficient and effective the knowledge has been used and the level of action based on it (Wilhelmij and Schmidt, 2000).

The KM literature clearly exposes that knowledge resources have been increasingly seen as an integral part of organizations’ value creating processes. In a similar vein, companies have become aware of the importance of intellectual capital of their own (Guthrie et al., 2003). Intellectual capital can be defined as ‘the sum of all the intellectual materials of a company’ – knowledge, information, intellectual property including trademarks, patents and licenses, experience and integrity, personnel competencies, collective brainpower, etc. – that is captured and leveraged to create value and that can be converted to wealth and profit (Stewart, 2001; Harrison and Sullivan, 2000; Bontis et al., 2000). Though there are a variety of different components that constitute intellectual capital, an
increasingly popular classification divides intellectual assets into three categories: human capital, structural capital and customer capital (Skyrme, 2002).

2.2. PERFORMANCE

The main objective of KM performance evaluation is to increase the effectiveness, efficiency and adaptability of KM efforts so as to add more value to the overall performance of the organization (Toften and Olsen, 2003).

Given the general rule about performance evaluation that performance improves through evaluation, it is reasonable to argue that measuring the outcomes and evaluating the contribution of KM practices are important to ensure the sustainability and success of KM efforts over time. Without assembling the link between desired outcomes and KM practices continuously and demonstrating tangible or quantifiable intangible results, it is not possible for the top management to keep on investing and for the workers to preserve their concentration and motivation (O’Dell and Grayson, 1998). Apparently, evaluation of KM-related organizational performance also shows to what extent the intellectual resources of a firm have been utilized (Firer and Williams, 2003; Marr et al., 2003) as well as the degree of the conversion of the organizational knowledge into improved performance (Kalling, 2003).

3. RESEARCH METHODOLOGY

3.1. SURVEY INSTRUMENT

Data were gathered via cross-sectional mail survey using a self-administered questionnaire that was essentially composed of questions related to KM practices and organizational performance. Respondents were asked to indicate the level of agreement based on five-point Likert scales ranging from 1 “strongly disagree” to 5 “strongly agree” on each of the items measuring eight aspects of KM practices, which include organizational structure (OS), technological infrastructure (TI), organizational culture (OC), intellectual capital (IC), knowledge generation (KG), knowledge codification and storage (KCS), knowledge transfer and sharing (KTS), and knowledge utilization (KU). All eight KM practices are essentially multi-item scales derived from a total of 28 items. While KTS and OS are measured by 7 and 5 items respectively, the following four KM practices including KCS, KU, KG and TI are all measured by 3 items. Both OC and IC are, however, measured through two items. Inter-item reliability measures of all 8 KM practices are well above the threshold value of 0.50 for surveys of exploratory nature as suggested by Nunnally (1978). Regarding the organizational performance related to KM, respondents were asked to rate to what extent KM practices have led to improvement on each of the following four organizational performance criteria over the last three years: overall organizational performance, usability of KM applications, overall employee performance and having a common sense of corporate mission. These organizational performance criteria related to KM practices were measured using five-point scales ranging from “definitely better” through “about the same” to “definitely worse” or “don’t know”. An index measure (PERF) composed of these four items captures the SME’s organizational performance in terms of KM practices ($\alpha = 0.87$).

The questions with regard to the study’s variables were derived from the extant literature and discussions based on semi-structured interviews with a number of senior managers from three SMEs. The questionnaire was pre-tested several times to ensure that the wording, format and sequencing of questions were appropriate. Occasional missing data on variables was handled by replacing them with the mean value. The percentage of missing data across all data was calculated to be relatively small.

3.2. THE SAMPLE

In line with small business research, this study adopted the number of employees as the base for the definition of SME. An SME is identified as one that employs fewer than 250 employees. The minimum of at least 10 employees was also chosen in order to exclude micro firms that would not be suitable for the purposes of this study. A self-administered questionnaire was distributed to 500 SMEs operating in textile industry within the greater metropolitan area of Istanbul in Turkey. For centuries being the largest city of Turkey, Istanbul has been undisputedly the main industrial and trade centre which accounts for nearly 75 per cent of total capital investment generating nearly 23 per cent of Turkish GDP (Berkoz and Eyuboglu, 2005). This sample was selected randomly from the database of Turkish Small Business Administration (KOSGEB). As of 2005, the KOSGEB database includes a total of 12,270 SMEs in the greater metropolitan area of Istanbul, which accounts for nearly 28% of all SMEs registered throughout Turkey. The sampling frame consists of 2,482 SMEs operating in the textile industry including textile mill products
and apparel (SIC codes 22 and 23), since it has been a leader in implementing progressive quality management practices in Turkey. The textile industry has also been the engine of economic growth and generates the largest volume of export revenues. Although one could argue that a focus on a single industry may make the results less generalizable, we ensured a high level of internal validity. Furthermore, within the textile industry there exist several different manufacturing environments and product types making the sample much more diverse than what could be expected for a homogenous sample.

It was requested that the questionnaire be completed by a senior officer/executive in charge of HRM and KM practices. The responses indicated that a majority of the respondents completing the questionnaire were in fact members of the top management. Of the 500 questionnaires posted, a total of 293 questionnaires were returned after one follow-up. Sixteen questionnaires were eliminated due to largely missing values. The overall response rate was thus 55% (277/500), which was considered satisfactory for subsequent analysis.

4. DATA ANALYSIS

The data analysis was conducted in two steps: First, we conducted a reliability analysis to check unidimensionality of each of the KM practices. Second, we measured the impact of each KM practice on the overall organizational performance using neural networks (NN) model.

4.1. UNIDIMENSIONALITY TESTS OF KM PRACTICES

The reliability of eight dimensions of knowledge management application was assessed by checking unidimensionality of each construct using two tools: Cronbach’s alpha and Dillon-Goldstein’s ρ (Chin, 1998). All of the Cronbach alpha values met the minimum threshold value of 0.70. Similarly, Dillon-Goldstein’s ρ analysis provides ρ values that are well above 0.70 for each construct supporting unidimensionality.

4.2. NEURAL NETWORK ANALYSIS

Recently, there has been a growing research attention in the development of artificial neural networks for solving a wide range of problems from different fields. Neural networks are distributed information processing systems composed of many simple computational elements interacting across weighted connections. Inspired by the architecture of the human brain, neural networks exhibit certain features such as the ability to learn complex patterns of information and generalize the learned information. Neural networks are simply parameterized non-linear functions that can be fitted to data for prediction purposes (Haykin, 1994).

The main appeal of neural networks is their flexibility in approximating a wide range of functional relationships between inputs and output. Indeed, sufficiently complex neural networks are able to approximate arbitrary functions arbitrarily well. One of the most interesting properties of neural networks is their ability to work and forecast even on the basis of incomplete, noisy and fuzzy data. In addition, they do not require a priori hypothesis and do not impose any functional form between inputs and output. For this reason, neural networks are quite practical to use in the cases where knowledge of the functional form relating inputs and output is lacking, or when a prior assumption about such a relationship should be avoided.

A neural network (NN) model used in this study has been developed in MATLAB to identify the most critical dimension of knowledge management for organizational performance of Turkish SMEs. To train and test the model, the data set was divided into two classes namely training and testing data sets. As a result, from the whole sample of 277 SMEs, 209 (75%) were selected randomly as training data, while 69 (25%) were selected as testing data.

The success of the NN models depends on properly selected parameters such as number of hidden nodes or neurons (HN) and layers, the nonlinear function used in the nodes, learning algorithm, initial weights of the inputs and layers, and the number of epoch the backpropagation model is iterated. The structure of the proposed NN model is shown in Figure 1, which consists of one input layer, one hidden layer and one output layer. Since there are 8 input variables denoting the KM practices, 8 nodes are placed at the input layer.
By changing the number of nodes (1 through 5) during the testing of the model, it is decided to use three nodes at the hidden layer, where the minimum mean square error (MSE) value occurs at this point for the training data set. As shown in Figure 2, the MSE values decrease for both the training and testing data sets when three nodes are used in the hidden layer. If there are more than three nodes at the hidden layer, it is possible to talk about the overtraining of the model. Thus, the network and its associated connection weights are set at this point and organizational performance according to NN model is estimated using these weights.

A hyperbolic tangent sigmoid and a linear transfer functions are used in the hidden and output layers, respectively. Levenberg-Marquardt backpropagation algorithm is applied to train the NN model since it trains faster, requires less memory and is not sensitive to learning and momentum rates. Also, the weights are initialized using the Nguyen-Widrow layer initialization method.

The performance of the NN Model is evaluated using MSE and R-square values. The best performance scores are estimated for the NN model with three nodes at the hidden layer. The MSE values for this model are calculated as 0.306 and 0.737 for training and testing data, respectively. The R-square value is found as 0.56, which could be
accepted as highly satisfactory due to imprecise nature of the data. Figure 3 shows the comparison between the NN estimations and exact values of the test data.

Figure 3: Comparison of NN model estimates with target values.

Assuming that NN model has an acceptable capability of predicting the organizational performance, the impact of KM dimension on firm performance may be evaluated and ranked based on the same model. This also provides managers with invaluable information in identifying which KM practices they should concentrate in order to have a better organizational performance. Table 1 shows the contribution of KM practices to the organizational performance in terms of the degree of their importance levels and their respective rankings. From the full set of 8 KM practices, knowledge transfer and sharing (0.147) appears as a leading factor, which is also in line with the existing KM literature. Technological infrastructure is the second most important criterion with the importance level of 0.140, while knowledge codification and storage is found to be the third important KM practice with the importance level of 0.139. In contrast, intellectual capital and knowledge generation are the least important KM practices in terms of their effects on organizational performance. This finding is not particularly surprising that KM practices have been primarily focused on knowledge transfer and sharing and technological infrastructure in SMEs.

Table 1: Importance of KM practices on the organizational performance

<table>
<thead>
<tr>
<th>KM practices</th>
<th>Importance level</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge transfer and sharing (KTS)</td>
<td>0.147</td>
<td>1</td>
</tr>
<tr>
<td>Technological infrastructure (TI)</td>
<td>0.140</td>
<td>2</td>
</tr>
<tr>
<td>Knowledge codification and storage (KCS)</td>
<td>0.139</td>
<td>3</td>
</tr>
<tr>
<td>Organizational culture (OC)</td>
<td>0.139</td>
<td>4</td>
</tr>
<tr>
<td>Knowledge utilization (KU)</td>
<td>0.135</td>
<td>5</td>
</tr>
<tr>
<td>Organizational structure (OS)</td>
<td>0.119</td>
<td>6</td>
</tr>
<tr>
<td>Intellectual capital (IC)</td>
<td>0.099</td>
<td>7</td>
</tr>
<tr>
<td>Knowledge generation (KG)</td>
<td>0.082</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION

This study has investigated the relationship between the implementation of KM practices and organizational performance. While it has been generally accepted that KM practices tend to enhance organizational performance, there is a paucity of empirical evidence to verify this link especially for SMEs. Our analysis provides a significant support to the contention that implementation of KM practices have a positive impact on the KM-related organizational performance of SMEs operating in Turkish textile industry.
Our findings indicate that the most important KM practices which influence organizational performance are namely knowledge transfer and sharing and technological infrastructure. KM practices and efforts for SMEs operating in textile industry have been primarily focused on sharing existing knowledge rather than knowledge generation. However, it should be recognized that the factors involving the KM processes are interrelated. That is, in order to improve knowledge sharing, the other factors should also be considered as a whole.

Nevertheless, the outcomes of this study are only valid for SMEs operating in textile industry of Turkey. As a further research area, a similar study can be conducted in different industries, and an industry wide comparison based on KM practices would enable to construct a KM map of the country. There is also a need to undertake comparative surveys examining the differences in KM implementation and their effect on organizational performance in SMEs and large-sized enterprises along with considering the other contextual and moderating factors affecting KM-performance link.

REFERENCES


A Metric for Rating the Effectiveness of Installed Industrial Automation Systems using an Adjusted Function Point Analysis

David W. Russell

Division of Engineering and Information Science
Penn State Great Valley
Malvern, PA 19355, USA

ABSTRACT

This paper describes the design of a factory automation performance and satisfaction metric that is based on variations to the function point analysis algorithm used in software engineering and the analytical hierarchy process methodology common to systems engineering. The paper asserts that a satisfaction rating can be obtained for completed automation projects based on this tautology. The metric uses five high-level functional ratings which are subdivided according to complexity criteria. The value of the metric is modified by calculating the effect of fourteen appropriate adjustment factors. A level of configurability is added by weighting these factors to suit any specific installation. The scores and configuration data are elicited by averaging results obtained from a multi-part survey administered to four or five company employees selected from a wide range of job titles and responsibilities. A simple implementation of the algorithm was completed, and an extension of the work will provide an Internet-based version and a diagnostic tool for company use.

1. INTRODUCTION

Automation projects are expensive and undertaken only after much deliberation especially in the current economic climate. Manufacturing companies increasingly must weigh technological progress against business expediency. Technologically, there are many tools and methodologies available to assist the system designer in the production of a wide range of manufacturing systems. In some cases a new construction project or machinery upgrade requires a corporation to requisition a custom (or packaged) manufacturing information system just in order to operate. Or perhaps the company is being compelled to comply with new regulations, or just seeking increased control over its bottom line. The design and implementation of such appear elsewhere in these proceedings. The author has been involved at both the technical and executive levels in factory systems for many years and his experience has been that few metrics exist that are expressly designed to assess the overall satisfaction of the company and the project success. The literature (for example [1], [2]) reports many project failures that involve IT and technology, and it may be safe to say that most “disasters” elude the media and never reach the public domain other than in court records!

This paper introduces a metric that was designed purely to give a corporation a fair and frank assessment of the success of an automation project across the landscape of the organization. The metric borrows some of the tenets of the Function Point Analysis [3] and Analytical Hierarchy Process [4] methodologies. The paper is organized as follows: Section 2 is a summary of those two methodologies. Section 3 shows the structure of the proposed Automation Metric and includes idealized data from a real factory. Section 4 outlines the data elicitation methodology. Section 5 briefly introduces the software, followed by summary, acknowledgements, and reference sections.

2. SUMMARY OF METHODOLOGIES

The following section summarizes relevant aspects of the Function Point Analysis and Analytical Hierarchical Process methodologies and how some of the features within the methodologies relate to the proposed metric.

---

1 Corresponding Author: Tel: +1 610-648-3233; E-mail: drussell@psu.edu
2.1 FUNCTION POINT ANALYSIS (FPA)

Function Point Analysis (FPA), created by Alan Albrecht in 1979, has become an ISO standard method of estimating software effort. It is based on producing a function point (FP) count from a detailed analysis of high-level functional requirements documents. Using statistical data from prior projects, FPA can estimate expected levels of effort, such as lines-of-code (LOC) or person-months depending on which implementation language (C, COBOL, etc.) is selected. To begin the process, analysts examine the requirements documents of the proposed system and assign values for the five major categories found in most IT systems - which are:

I. The number of outputs expected in the final system – for example; displays, reports, voltages.
II. The number of inquiries expected in the final system – this includes keyboarding, card swipes.
III. The number of input types – this includes keyboards, sensors, biometric data, scans.
IV. The expected number of files, both permanent and temporary, to be accessed.
V. The number of interfaces the proposed system must make with other systems.

Each value is then divided according to its apparent or expected complexity as simple, average, or complex. For example; if one system interface is into an existing SAP™ implementation, it would be classified as complex, whereas, a keyboard entry requirement would be added to the simple counter etc. Table 1 illustrates the initial phase of the method with data from a fictitious system. In this example, row 3 codifies that the specification has called for 100 inputs, of which 20 are considered simple, 50 are average, and 30 are assessed to be difficult. The application complexity matrix for that category is therefore \{20, 50, 30\} – and so on for all five categories.

Table 1: Application Complexity Matrix ACM

<table>
<thead>
<tr>
<th>Row</th>
<th>FUNCTION</th>
<th>DATA</th>
<th>COMPLEXITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nd</td>
<td>Simple</td>
</tr>
<tr>
<td>1</td>
<td>I: OUTPUTS</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>II: INQUIRIES</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>III: INPUTS</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>IV: FILES</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>V: INTERFACES</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

Data from prior projects is used to rank the expected difficulty in producing each component in the design. Table 2 shows the typical complexity factor matrix used in many software projects.

Table 2: Complexity Coefficient Matrix CM

<table>
<thead>
<tr>
<th>Row</th>
<th>FUNCTION</th>
<th>COMPLEXITY COEFFICIENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Simple</td>
</tr>
<tr>
<td>1</td>
<td>I: OUTPUTS</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>II: INQUIRIES</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>III: INPUTS</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>IV: FILES</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>V: INTERFACES</td>
<td>5</td>
</tr>
</tbody>
</table>

The value of the number of “unadjusted” function points (UFP) is obtained by multiplying ACM by CM and summing terms across the rows which, for this data set, yields a value of 2050. Because of the domain complexities of most systems, this initial estimate is naive and needs to be adjusted to take into account various influence factors [4] for the project as a whole. The actual Function Point count (FP) is calculated by multiplying the Unadjusted Function Point count by a Value Adjustment Factor (VAF). In FPA, the VAF coefficient is calculated from the individual weighting of fourteen factors such as reusability, level of distributed computing, software integration, etc. Because this factor is used in the metric below, it is not discussed in any detail here. The adjusted function point value, \( FP = UFP \times VAF \) can then be used to estimate [4] the project’s expected duration, level-of-effort and the number of lines-of-code in any particular language using a table or graph. For the values given in Table 1, with an example value adjustment factor (VAF) of 1.26, the adjusted FP
value would be re-calculated to 2583. There are several online function point calculator tools [6] available. Figure 1 is a screen capture from http://ivs.cs.uni-magdeburg.de/sw-eng/us/java/fp/. This shows that the estimated level-of-effort for the example project with 2583 function points would be 347 person months. This represents a large project that would require a team of analysts over several years.

![Figure 1: Output from a Function Point Calculator](http://ivs.cs.uni-magdeburg.de/sw-eng/us/java/fp/)

2.2 ANALYTICAL HIERARCHY PROCESS (AHP)

Because most problems have solution options, each one with its own aspects of excellence, complexity, and cost, it is imperative to be able to analytically review and evaluate options based on multiple criteria. The AHP method [6] first establishes a hierarchical scale of goals and sub-goals. Each goal is rated on a scale of 1 to 9, with the value “1” conveying equal importance and “9” indicating a completely more important feature. For example, a design with four goals could be evaluated as shown below in Table 3. The “9” at table position \{2, 3\} indicates that Goal C is very much more important than goal B. For reference, the lower diagonal matrix is just the reciprocal of the upper.

<table>
<thead>
<tr>
<th>Table 3: AHP Level Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal A</td>
</tr>
<tr>
<td>Goal A</td>
</tr>
<tr>
<td>Goal B</td>
</tr>
<tr>
<td>Goal C</td>
</tr>
<tr>
<td>Goal D</td>
</tr>
</tbody>
</table>

The system goals are now sub-divided into criteria which are weighted mathematically using an Eigen vector analysis to obtain the dominant values. Each criterion is assigned a relative weight for each goal, normalized, and multiplied across the level matrix to form composite weights. The criterion with the highest composite figure is an indicator of the preferred option to take in the design. Table 4 shows an example of this where the composite weight of criteria 1 is 0.238 \{=0.507x 0.2 + 0.359x0.3 + 0.094x0.1 + 0.039x0.5 = 0.238\}.

<table>
<thead>
<tr>
<th>Table 4: AHP Goal –v - Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
</tr>
<tr>
<td>Criteria 1</td>
</tr>
<tr>
<td>Criteria 2</td>
</tr>
<tr>
<td>Criteria 3</td>
</tr>
<tr>
<td>Criteria 4</td>
</tr>
</tbody>
</table>
Table 4 would indicate that criteria 2 is the most significant of the four and is shown shaded. This is by no means intended as a complete description of the AHP method but rather as a simple overview of this decision support methodology, and the interested reader is directed to the literature.

2.3 RELATION TO PROPOSED METRIC

Both of these fairly well known methods provide useful means in the estimation and selection of resources but are both customarily applied before the onset of a project. FPA measures expected complexity and AHP rates viable options based on system variances. The proposed automation metric (AM) described below uses a similar tautology to FPA as indicated by the scoring the various facets of a completed project which are based on the input of actors from within and possibly outside the company. In addition, a weighting and ranking method similar to that found in AHP is used to apply some level of discrimination between factory to factory variations, and to compensate for human inputs across a wide employee spectrum and their corresponding value systems. For example, an operator may value ease-of-operation while an executive might focus more on increased throughput.

3. THE AUTOMATION METRIC (AM)

The proposed metric is designed to give a measure of overall satisfaction for an automation project in a factory setting. The metric is an attempt to objectively evaluate the true worth of an automation project.

3.1 THE UNMODIFIED AUTOMATION METRIC (UAM)

In order to accomplish preliminary data, the so called AM metric utilizes five high-level automation function (AMF) values, each of which is then distributed according to perceived and actual complexity. The essential difference from the FP methodology is that these values are elicited from a survey given to four or five employees ranging in rank from company executive to shop floor operators. The complexity coefficients ($C_k$) are only estimates based on vendor and user opinions. Table 5 contains data that illustrates the calculation of the UAM or unadjusted automation metric.

Table 5: Computation of Unadjusted Automation Metric (UAM)

<table>
<thead>
<tr>
<th>AUTOMATION FUNCTIONS (AMF) – based on responses to a questionnaire with values between 10 and 100.</th>
<th>COMPLEXITY DISTRIBUTION - based on the complexity of the installation – “CM” values</th>
<th>ROW TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Row</strong></td>
<td><strong>Data Values</strong></td>
<td><strong>Simple</strong></td>
</tr>
<tr>
<td>1</td>
<td>COST (EFFECTIVENESS) What has been the impact of the automation?</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>OPERATION Do operators find the system easy or hard to use?</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>TECHNOLOGY How does the system rate as far as reliability, accuracy and availability?</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>PRODUCTIVITY Did the automation improve efficiency?</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>INTEGRATION How well does the system work with prior business functions?</td>
<td>65</td>
</tr>
<tr>
<td><strong>UAM TOTAL/MAX</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the data given in Table 5, the value of UAM is 2235 out of a maximum possible score of 3700. Taking this as a percentage gives the simplest measure of Satisfaction Level (SL) as ~60% (= 2235/3700*100).
3.2 AUTOMATION ADJUSTMENT FACTORS (AAF)

Because each factory has very different technical performance measures (TPM) and structures, it is again expedient to modify the raw UAM figure using (coincidentally) fourteen Automation Adjustment Factors [AAF] in order to yield a better value. Each item in the table is given a value between 1 and 9 as in AHP. An AAF that performs very well is highly rated and given a score of 9/9, a so-so rating 5/9 and a poor rating of 1/9. Equation 1 is used to limit the VAF influence to ±35% as in FPA – note that the maximum AAF total possible is 126 (=9x14) should all factors be given 9/9.

\[
VAF = 0.65 + 0.70 \times \left( \sum_{i=1,14} \frac{\{AAF_i\}}{126} \right)
\]  

Equation 1

Table 6: Automation Adjustment Factors (AAF) with Example Values

<table>
<thead>
<tr>
<th>#</th>
<th>Adjustment Factors</th>
<th>Raw AAF</th>
<th>Reasons and Comments for the Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Communications</td>
<td>7/9</td>
<td>Are communications easy to establish and maintain?</td>
</tr>
<tr>
<td>2</td>
<td>Web Enablement</td>
<td>3/9</td>
<td>Is information to be Web accessible?</td>
</tr>
<tr>
<td>3</td>
<td>Productivity</td>
<td>9/9</td>
<td>Has the product increased production?</td>
</tr>
<tr>
<td>4</td>
<td>Agility</td>
<td>9/9</td>
<td>Is it easy to alter job settings often?</td>
</tr>
<tr>
<td>5</td>
<td>Transaction Rate</td>
<td>9/9</td>
<td>Is the product in step with the process?</td>
</tr>
<tr>
<td>6</td>
<td>On line Data Entry</td>
<td>7/9</td>
<td>Do operators enter data from the shop floor?</td>
</tr>
<tr>
<td>7</td>
<td>Efficiency</td>
<td>5/9</td>
<td>Is this an easy and efficient process?</td>
</tr>
<tr>
<td>8</td>
<td>User Training</td>
<td>7/9</td>
<td>Was it easy to train operators?</td>
</tr>
<tr>
<td>9</td>
<td>Complexity of System</td>
<td>5/9</td>
<td>How difficult is it to keep the system running?</td>
</tr>
<tr>
<td>10</td>
<td>User Experience</td>
<td>1/9</td>
<td>How much prior experience was necessary?</td>
</tr>
<tr>
<td>11</td>
<td>Ease of Installation</td>
<td>7/9</td>
<td>How easy was it to install and bring on line?</td>
</tr>
<tr>
<td>12</td>
<td>Shop Floor Effects</td>
<td>3/9</td>
<td>Is it easy to apply patches and program fixes?</td>
</tr>
<tr>
<td>13</td>
<td>Multiple Sites</td>
<td>1/9</td>
<td>Is the product to be installed elsewhere?</td>
</tr>
<tr>
<td>14</td>
<td>Cost Efficiency</td>
<td>7/9</td>
<td>Has the automation affected the bottom line?</td>
</tr>
</tbody>
</table>

For the data given in Table 6, using Equation 1, the value of VAF is 1.09 (=0.65 + 0.7*80/126). This would then make the adjusted automation point score increase from 2235 to 2436 and seemingly increase the satisfaction percentage from 60% to ~66% ... but this is not so because the maximum possible value of AM is now 4995 (=1.35 * 3700). In actuality, the satisfaction percentage is reduced to ~49% (=2346/4995). Influence factors adjust the automation metric and, therefore, the satisfaction levels to better reflect how satisfied the company is with the results of its automation project.

3.3 CONFIGURATION BY WEIGHTING

In the previous section, the adjustment factors were all considered equally significant in any factory. To incorporate variances in their significance within any particular factory, each adjustment factor is also assigned a weight. This is to avoid an AAF from being over-rated when it is really unimportant. This will further modify the AM value, but not the maximum value attainable as this remains at 1.35. Table 7 is an extension of Table 6 and shows how the company might weight each of the adjustment factors to match their particular installation.

For the data given, the adjustment factor total is now 36.77 out of the new maximum of 70.2 (=9 x 7.8). Using Equation 2, the new value of VAF is 1.01 (= 0.65 + 0.7 * 36.77/70.2), giving a new value of AM of 2257, which in turn translates to a reduced satisfaction level of ~45%.

\[
VAF = 0.65 + 0.70 \times \left( \sum_{i=1,14} \frac{w_i \times \{AAF_i\}}{\sum_{i=1,14} w_i} \right)
\]  

Equation 2

The rationale for this weighting is simple. For example, the AAF6 feature (on line data entry) – shown shaded in Table 6—might work very well as demonstrated by its high score of 7 out of 9, but the company may have subsequently decided to limit data entry on the shop floor and consequently rank this feature as being “neutral” in the rubric shown in Table 7 where item 6 has been assigned a weight of only 0.5. This weighting avoids over-rewarding “something insignificant that works well” as opposed to “something important that works well.” Without such weighting the metric could be skewed upwards by good performance in ancillary features. The data in Table 7 represents a factory with long repetitive job runs and little operator intervention.
Table 7: Weighted Adjustment Factors with Example Values

<table>
<thead>
<tr>
<th>#</th>
<th>FACTOR</th>
<th>Configuration Rubric</th>
<th>Assigned Weight, $w_i$</th>
<th>Raw AAF$_i$</th>
<th>Weighted AAF$^i*wi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Communications</td>
<td>Neutral</td>
<td>0.5</td>
<td>7/9</td>
<td>3.5</td>
</tr>
<tr>
<td>2</td>
<td>Web Enablement</td>
<td>Neutral</td>
<td>0.5</td>
<td>3/9</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>Productivity</td>
<td>Unimportant</td>
<td>0.1</td>
<td>9/9</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>Agility</td>
<td>Unimportant</td>
<td>0.1</td>
<td>9/9</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>Transaction Rate</td>
<td>Unimportant</td>
<td>0.1</td>
<td>9/9</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>On-line Data Entry</td>
<td>Neutral</td>
<td>0.5</td>
<td>7/9</td>
<td>3.5</td>
</tr>
<tr>
<td>7</td>
<td>Efficiency</td>
<td>Neutral</td>
<td>0.5</td>
<td>5/9</td>
<td>2.5</td>
</tr>
<tr>
<td>8</td>
<td>User Training</td>
<td>Important</td>
<td>1.0</td>
<td>7/9</td>
<td>7.0</td>
</tr>
<tr>
<td>9</td>
<td>Complexity of System</td>
<td>Important</td>
<td>1.0</td>
<td>5/9</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>User Experience</td>
<td>Important</td>
<td>1.0</td>
<td>1/9</td>
<td>1.0</td>
</tr>
<tr>
<td>11</td>
<td>Ease of Installation</td>
<td>Important</td>
<td>1.0</td>
<td>7/9</td>
<td>7.0</td>
</tr>
<tr>
<td>12</td>
<td>Shop Floor Effects</td>
<td>Neutral</td>
<td>0.5</td>
<td>3/9</td>
<td>1.5</td>
</tr>
<tr>
<td>13</td>
<td>Multiple Sites</td>
<td>Neutral</td>
<td>0.5</td>
<td>1/9</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>Cost Efficiency</td>
<td>Neutral</td>
<td>0.5</td>
<td>7/9</td>
<td>3.5</td>
</tr>
</tbody>
</table>

TOTALS 7.8 80/126 = 0.635 36.77/70.2=0.52

Value Added Factor = 0.65 + 0.7*sum/max= 1.09 1.01

3.4 COMPARISON OF VALUES

The effect of each refinement is shown in Table 8 for the same original data values indicating that the initial data values were rather generous; whereas, when the adjustments are factored in, the real satisfaction level percentage was under 50%.

Table 8: Effect of Weighted Adjustment Factors

<table>
<thead>
<tr>
<th>Metric</th>
<th>Automation Metric Value (AM)</th>
<th>Maximum Metric Value (AMx)</th>
<th>Satisfaction Level AM/AMx (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified</td>
<td>2235</td>
<td>3700</td>
<td>60.4%</td>
</tr>
<tr>
<td>Adjusted</td>
<td>2436</td>
<td>4995</td>
<td>48.8%</td>
</tr>
<tr>
<td>Weighted</td>
<td>2257</td>
<td>4995</td>
<td>45.2%</td>
</tr>
</tbody>
</table>

4. DATA ELICITATION

In order to populate the tables with meaningful data, a six-part questionnaire was designed. Because data obtained from company personnel (the actors) is qualitative as well as quantitative, careful data collection by an analyst is imperative. For example, assessments proffered by a disgruntled, or soon to be terminated, employee must not be given the same weight as the opinions from more objective sources. Qualitative and quantitative feedback is solicited from staff of varying levels and depths of interface with the factory system. The questionnaire is in six parts with each sub-section contributing to the data in Tables 5, 6, and 7.

4.1 COMPANY AND LEGAL

This section is to be completed by a company executive and may include attachments such as non-disclosure forms and permits. Even though the metric can be configured to be completely anonymous, this proved to be the most difficult part of the questionnaire due to the oftentimes proprietary nature of the production processes at a plant and the reluctance of executives to go on record or to allow employees access to a forum where grievances or illegalities might be aired. Without this buy in, the other parts of the questionnaire are impossible. A common excuse has been that the technical and operational staff does not understand the complexities of capital item purchasing with its own set of tax implications, or the disjoint between equipment and payroll funding. A common complaint is why the millions spent on equipment and systems could not have been used for salary increases.

As an illustration, an extract of Part A of the questionnaire is shown as Table 9.
Table 9: Section of Questionnaire Part A

<table>
<thead>
<tr>
<th>REF</th>
<th>QUESTION</th>
<th>INFORMATION ENGINEERING</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 01</td>
<td>Company Trading Name and Address</td>
<td>Get formal company name and location of headquarters.</td>
</tr>
<tr>
<td>A 02</td>
<td>Can we publish your company name and location – can we publish results</td>
<td>Identify persons who can sign a legal release form – e.g. a non-disclosure form?</td>
</tr>
<tr>
<td>A11</td>
<td>Any future plans in regard to automation?</td>
<td>Probing plans for the Factory of the Future</td>
</tr>
<tr>
<td>A12</td>
<td>Would you like a copy of the final report?</td>
<td>Ascertains reporting paths</td>
</tr>
</tbody>
</table>

4.2 AMF VALUES

The second part of the questionnaire focuses on obtaining overall AMF values and estimating the perceived complexity ratings for each; this part is administered to all interviewees. Data are averaged and the values inserted into Table 6. This yields a broad overview of the company opinions in regard to the success of an automation project and the satisfaction level was calculated, for the given data, to be around 60%.

4.3 ADJUSTMENT FACTORS AND WEIGHTS

This part is used to establish adjustment scores and weights and is administered to all non-executive actors. The averaged values obtained are inserted in Tables 6 and 7. This is a broad overview of the climate in the automation landscape and determines what features of the systems are really important.

4.4 TECHNICAL ISSUES

This part is used to gain insights from the on site IT and Engineering staff who may have installed or integrated the system, or who must now maintain it and who have in-depth knowledge of technical issues. This helps build Tables 5 and 7. The target actors are engineers and consultants, who are always ready and willing to offer opinions!

4.5 OPERATIONS MANAGEMENT

This focuses on local expertise and in-depth knowledge of the mechatronic systems in place by those who must maintain the production line components and resolve schedule and operator issues daily. This relates to the values in Tables 6 and 7.

4.6 PRODUCTION OPERATORS

This part solicits daily occurrences, perceptions, and experiences of the operators and helps build Table 6.

4.7 ADMINISTERING THE QUESTIONNAIRE

Table 10 describes the strategy in regard to which parts of the questionnaire will be administered to which actors. The lowercase “yes” implies that input from that group should not be considered a primary source.

Table 10: Interview Strategy

<table>
<thead>
<tr>
<th>ACTOR</th>
<th>ROLE</th>
<th>QUESTIONNAIRE see SECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXEC – Company Executive</td>
<td>Provides the corporate overview of the automated system</td>
<td>YES NO</td>
</tr>
<tr>
<td>TECH - Member of Technical Staff</td>
<td>Understands the computational intricacies of the automation systems</td>
<td>YES NO</td>
</tr>
<tr>
<td>OPMGT- Operations Manager</td>
<td>Provides the day-to-day information in regard to how easy the system is to work</td>
<td>YES</td>
</tr>
<tr>
<td>OPSTAFF-Shop Floor Operator</td>
<td>Provides the grass-roots opinion of the reliability and usefulness of the system</td>
<td>YES</td>
</tr>
</tbody>
</table>
5 SOFTWARE INTEGRATION WITH THE QUESTIONNAIRE

The AM algorithm was coded into a simple software tool that allows any number of major phyla (AMFs) and any number of weighted adjustment factors (AAFs) to be used. As more AMFs are configured, the metric simply gives larger values to be divided by larger perfect score values. The case tool questionnaire contains anecdotal responses to sets of questions from which numeric data is extracted for processing by the metric. Textual data is simply aggregated into an opinion box for reporting usage.

6 SUMMARY

Project design and evaluation [7] is the bedrock of systems and software engineering. Before the go ahead for any moderately large project is given, it is customary to produce conceptual documents, including as detailed a system specification as is possible before development begins. Tools such as function point analysis and AHP have proved very useful in estimating resources and decisions that must be made in regard to options.

Before a new project can be started it is the author’s opinion that reference to how prior projects were accomplished should be made and especially so before selecting an outside vendor or integrator. Anecdotal and financial references are only partly useful as it is axiomatic that the vendor will provide only the names of satisfied customers. What is needed is a quantitative metric that can provide an unbiased evaluation of completed work. In the manufacturing enterprise, where in-house technical expertise and R&D may be limited, necessary automation projects often mandate the services of an outside vendor, so having data about prior work is essential.

The paper has postulated a simple tool designed to calculate automation points is a measure of automation efficiency and true satisfaction level. The AM method is a metaphor of the Function Point Analysis and Analytical Hierarchy Process methodologies but with major differences. The first is that AM data is obtained in retrospect from completed systems in existing installations. Secondly, input about the usability and acceptance of a system is solicited from the opinions of executives, users, and technical staff to assure a more balanced viewpoint than is possible from system analysts alone. Thirdly, the weighting of adjustment factors attempts to provide a level of configuration in any particular plant. Finally, the AM value obtained for any given site is intended to assess the real worth or value of an automation project to a company.

The project is still active with company visits being planned in several countries of the world. An extension to the project will make the survey Web-based and access to the software granted on a licensing basis. By incorporating some of the qualitative data, it is envisioned that useful “post-mortem” reports can be generated and used as part of a roadmap to improvements in future projects.

ACKNOWLEDGEMENTS

The author acknowledges the contribution of The Pennsylvania State University for its sabbatical support of the project. Some parts of this paper also appear in reference [8] and are included with permission. Materials regarding AHP are taken from a term paper written by, and used with the permission of, Mike Zsilavetz from the Lockheed Martin Corporation who is a (post)graduate student at Penn State Great Valley.

REFERENCES

CAN Protocol in diffuse controlling system based on B&R hardware

J. Swider, G. Kost, G. Wszolek, D. Reclik

Institute of Engineering Processes Automation and Integrated Manufacturing Systems, Silesian University of Technology, Gliwice, POLAND

ABSTRACT

The scope of the paper is the description of the distributed control system in B&R Company based on Programmable Logical Controller (PLC) Power Panel 220. The discussed control system includes the Central Processing Unit (CPU) integrated with the touch screen and distributed input and output system. The entire process of data exchange between the CPU of the PLC and the external input-output modules runs along the internal Controlled Area Network (CAN). The distributed control system operated by B&R is a modular solution. The connection of successive input and output modules is possible thanks to the application of the docking station that serves the CAN data rail protocol. Another issue discussed in the paper is the connection of numerically controlled axes to the control system. The addition of a servo-operator is possible by means of the same CAN protocol that is used in the communication between the CPU and the inputs and outputs. The presented solution has multiple large-scale applications involving, for example, the control of 6- and more axis robots, numerical machine tools, process technologies control, remote data acquisition. The paper is substantiated by a real application of the distributed control system implemented in B&R Company in the form of shelves store control.

1. INTRODUCTION

Continuous progress in the fields of industrial electronics and automatic control has contributed to a wide application of advanced technologies of information dispatch in almost all control and adjustment applications [11,……,14]. Network solutions applied in remote collection and dissemination of data have facilitated the expansion of control systems, without the need to use a bigger number of PLCs, as, thanks to the utilization of an industrial network (Profibus, CAN, Ethernet), it is possible to connect the input and output modules to one PLC even if they are located at the distance of several meters away [1, 2, 5, 6].

In the face of wide range of applications (the automotive industry, diagnostics of ships, connections to anti block system (ABS), Electronic Brakeforce Distribution (EBD), etc.), the authors of this paper have decided to utilize the CAN protocol in a distributed control system, in view of its many advantages, which simplify the distributed control of process technologies and ensure a high level of safety. Because the CAN protocol is based on dispatch and receiving messages, controllers and other equipment that transmit the message do not have to know the recipient’s address. As the headline of the message contains the information concerning the addressee, if it arrives to the target unit (often referred to as: CAN node), it will be read out, otherwise, it will be ignored. The solution based on messages rather than addresses makes it possible to receive information simultaneously by many recipients, which increases the speed of information flow in comparison with standard information systems based on sequences through successive network nodes. For example, measurements required for several PLCs may be broadcast by means of the CAN protocol. Thus, the used transmission band is minimized. Another attribute of the CAN protocol limiting the transmission band is called Remote Transmit Request and makes it possible for the system node to send the message through other nodes. Such solution is particularly useful in diagnostic programs. Instead of ordering other stations periodic forwarding of the information on the condition of the system, it is possible to require the information from the nodes that are of lesser importance to its proper operation only if a fault is detected. Another advantage of the CAN main is fault confinement. The main prevents the whole system to be stopped if a problem
occurs only at one node. CAN nodes may perform self-diagnostics and, depending on the gravity of the detected fault (for example: the fault is permanent or temporary) switch to one of the three possible operation manners, including total stoppage. This feature prevents the faulty/damaged node from blocking the communication in the network by continuous notification about the faults, thanks to which the remaining part of the process may run without disturbances [1, 5, 6, 7, and 8].

2. CAN PROTOCOL IN DISTRIBUTED CONTROL SYSTEMS USED IN B&R COMPANY

The CAN solution applied by the authors of the paper was implemented in the subsections of B&R Company, which became an industry leader in innovation and technology and will continue to grow in the coming years and will further strengthen their leading position in the international market.

As the entire project was run on Power Panel 220 controller (with active touch screen) the PLC had to be equipped with the CAN interface to enable the construction of an integrated network structure. Factory made PP220 controllers [4, 5, 6, 7, 8] (also see Fig.1.), do not have the function of serving industrial network, but have an option of adding expansion bus. Such solution enhances the universal nature of PLCs applications, as it is possible to add two cards that extend the operation to industrial networks, including CAN (depending on the type of the controller it may only be one card of network operation). The described project used IF771 card [8, 9] (Fig.1.) for serving the internal CAN with an option of connecting other producers’ nodes (CAN items) by means of CANOPEN protocol emulation [9].

![Fig.1 CAN structure including the added subsections/subsystems. 1. B&R Power Panel 220 PLC, 2- IF771 card for serving CAN, 3- I/O modules docking station of CX470 CAN type, 4- CM211 digital and analogue input-output module, 5- DH439 digital input module, 6- ACOPOS 1016 B&R servo-operator with 8MSA3M engine, 7- expansion bus for servo-operator capacity to include the service of 8AC110.60.2 communication](image)

The PLC equipped in such manner constitutes the first node of an industrial network. As every automatic control system, apart from control units, must also contain operational/executive units, they must be added in the form of the digital and analogue input-output modules to the established one node CAN, as well as the CAN docking stations.
EX470 docking stations (see Fig.1) were used in the project, making up the successive network nodes (NODE#2 and NODE#3). In the next step, CM211 digital and analogue input-output module \[ 6, 8, 9 \] (Fig 1) was connected to the second node, which constituted the docking station. Thus, it was possible to connect the pressure valves and pneumatic valves to the system, and, at the same time, to control the operation of the actuators. As every automatic control system may not be grounded on “trust” in its algorithm, it was necessary to adjoin a proper system of sensors of the position of pneumatic pistons [3] (including the measuring encoder). To achieve this, DI439 digital input modules \[ 8, 9 \] (Fig.1) were connected to nodes 2 and 3 (docking stations) of the constructed CAN. Such configuration facilitated the achievement of the majority of the assumed tasks. Because the necessity of very exact positioning is often one of the major tasks of industrial automatic control systems (as in the case of the discussed application), a numerically controlled axis was also added to the system [3]. The servo-operator used in the project was the B&R one with intelligent ACOPOS 1016 traffic controller [8, 9] (Fig.1). As it was not equipped with the CAN interface, it was necessary to furnish it with 8AC110.60-2 expansion bus [8, 9], (Fig.1). The extended servo-operator, together with the CAN made up another, fourth node of the industrial network (NODE#4). Thus, a functional and autonomous automatic structure was established on the bases of distributed control along the CAN data rail. The layout of all elements of the system is shown in Fig.1. It should be emphasized that the application of a network solution requires every new node to have a constant number, the value of which is designated by two mechanical set-ups (approximate and exact adjustment knob), setting the number in the hexadecimal code (eg. set point 0 and F correspond to number 16) [8, 9].

3. STRUCTURE OF THE KINEMATICAL CHAIN OF HIGH STORAGE MANIPULATOR

The network solution proposed by the authors of this paper constitutes the foundations of the distributed control system which was used for continuous supervision and operational control of high storage. To enable the operation of store indexes involving the collection of details from the production process, forwarding them to the index indicated by the control system, placing them at the right storage indexes and dispatching, upon request, specific products to the collection point, an integrated manipulator with at least 4 degrees of freedom was employed. The following modules are contained in the kinematical chain of the electro-pneumatic manipulator serving the high storage (Fig.2):

- horizontal movement module
- vertical movement module
- gripping arm turn module
- gripping arm reach module
- gripping devices changer together with the devices

Such division is an outcome of the variety of the kinematical couples making up the kinematical chain of the manipulator, as well as of the variety of functions and tasks that they must achieve.

![Fig.2. Diagram of the kinematical chain of a 4-axis electro-pneumatic manipulator for high storage](image)

To drive the horizontal axis integrated with the high storage manipulator, LINTRA M46140/M/25000 bypass piston-free pneumatic servo-operator was used, manufactured by Norgren Herion- see Fig.3.
The piston-free engine was additionally equipped with bracing truck guide bars securing a significant stiffness of the system, improving the accuracy of positioning which is so essential while collecting elements from the store and changing the gripping devices. The horizontal drive is supplied through E/P VP1006BJ101A00 proportional valves produced by Norgren Herion – see Fig.4.

For the measurements of a current position of the horizontal manipulator, Wobit incremental encoder with integrated linear transducer was used, with the measuring capacity of 2500 mm.

In view of the required accuracy of positioning, the vertical axis drive contains M/49132A/BS/052/750 encased cross-helical gear with no clearance manufactured by Norgren Herion. On the bases of the steady operational parameters, the selected drive is B&R 8MSA3M.E0-I6-Rev.D1 controlled by ACOPOS 1016 amplifier shown in Fig.5.

The gripping arm of the manipulator consists of two modules: the reach module and the turn module. The gripping devices changer module is contained in the hand of the manipulator. The gripping arm integrated with the high storage pneumatic manipulator is shown in Fig.6.

The pneumatic servo-operator that turns the arm is M/60272/M/180 Norgren Herion with the operational range of 180°. To secure an option of manipulating the objects placed on the shelves of the store, the gripping arm module is equipped with RM/8021/M/250 pneumatic by-pass piston servo-operator (also produced by Norgren Herion) with the integrated QM/8021 bracing guide bars- see Fig.7.
The gripping arm of the manipulator is also equipped with automatic changer – see Fig.8- constructed on the bases of Norgren Herion mini servo-operator [10], which, by means of an additional element, switches the blockout joint (Norgren Herion).

4. IMPLEMENTATION OF THE OPERATIONAL ALGORITHM FOR THE CONTROL SYSTEM BASED ON B&R DISTRIBUTED SYSTEM

The first step made in the course of creating a parametric program that controls the operation of high storage was the configuration of the equipment. Accordingly, a new project was set up in B&R’s Automation Studio software environment, and, first and foremost, a decision undertaken concerning an applicable type of PLC. The selected PLC is PP220.1505-B5. The Automation Studio, on the grounds of the chosen PLC type, selected the right version of the operational system. Once the configuration of the basic project was finished, it was possible to move on to the main screen of the environment- see Fig.10.
In the left part of the screen there is a hardware configuration tree. The right part of the screen contains detailed characteristics of the object indicated by the tree. Due to the indication of the master object in the form specifying the name of PLC, the right part of the screen displays a tree of the installed components (libraries, visualization objects and program tasks). To introduce the controller of servo-operator ACOPOS 1016 to the hardware tree, as well as the analogue and digital input and output modules, it was necessary to define the expansion bus to be installed in the PLC and other successive components connected to CPU through CAN. After defining the docking stations and the input and output modules, in the fourth node of the network the controller of ACOPOS servo-operator was defined and an engine cooperating with the servo-operator (object 8V1016.50.2).

Upon forwarding the created project to the PLC it was possible to start-up the environment testing the servo-operators. The test mode enables the assumption of control over the operation of the servo-operator by ON-LINE computer, thus, the action of the PLC is replaced by the servo-operator. In the course of starting up the vertical axis drive of the pneumatic manipulator which is integrated with the high store, the “Test” environment was used by designating the optimal values of the displacement and velocity errors reinforcement coefficients. The left upper part of the test module screen (Fig.11) shows a set of instructions contained in the operational structure of the numerically controlled axis. The instructions available in the window (1) may be extended by other ones accessible from the programming language level.

In the right lower part of the screen (2) the parameters of the states of the servo-operators are displayed (structure “p_ax_dat.monitor). Above the instruction screen (1) there is a signal of the readiness of the servo-operator and the emergency STOP button for the servo-operator engine. Above the states readouts (2) there is a signal of the engine supply, turning on the positioning adjustment, states readouts, drives and servo-amplifier errors sensor. In the central part of the screen there is an access window to the servo-drive structure (3), and, at its right side (3) the editor of the parameters values concerning the selected substructure (4). To facilitate the tracking of the parameters values, the oscilloscope window is displayed at the bottom of the screen (5). In view of a specific nature of single-phase B&R drives, it was necessary to switch off the standard monitor of the three-phase drive and to introduce a default change of the effective value of the supply servo-amplifier. To change the default parameters the service channel is used. The service instructions are sent in three stages. First of all, the number of the service instruction should be set-up. After that, a new value of the substructure entered (data_tex). The execution of these activities results in a new value of an indicated service parameter to be saved in the buffer. The change of the service set-ups shall be active until the Test environment is abandoned.

The initiation of the procedure of adjusting the servo operator entails the change of the velocity error reinforcement coefficient. In the first phase of the adjustment (loading the engine shaft), this coefficient kv =0. The
adjustment of the displacement error requires the start-up of the servo-drive and tracking the displacement variable and the velocity. The adjustment of the displacement reinforcement coefficient involves a gradual increase of the value of kv, until self-excited vibrations of the engine shaft appear, so the final value of kv was assumed 12. This value is a safe boundary value of the coefficient, above which considerable increase in the velocity oscillations are observed. After the adjustment procedure for the displacement reinforcement coefficient is finished, the velocity control parameters are adjusted, on the grounds of the proportional dependence between the moment and the current intensity in the engine winding and by means of Fourier transform (FFT) which is available as an embedded function in the Automation Studio environment.

5. SUMMARY

The CAN solution utilized by the authors of the paper constitutes a perfect example of a distributed control system. A single B &R PLC used in the system enables the connection of 255 nodes of basic CAN, so the control system may be extended to a very large degree. Such solution may provide simultaneous, safe and functional control of many different subsystems of industrial automation. Thus, it may be concluded that such solution used for big applications enables the elimination of a significant number of logical PLCs in favour of just one controller with the implemented CAN service, leading to the reduction of the costs of the control system of the entire application, and improved clarity of connections (cabling), with the parallel easiness in identifying the input and output signals. The use of distributed control also makes it possible to avoid measurement errors caused by the drop in the measured voltage values (analogue measurements) along the signal wires, if the module with the A/C converter is placed in the nearest vicinity of the object, the voltage signal of which is to be measured. The information sent to the master PLC is converted into a digital form, which is not sensitive to the phenomenon of voltage drop (while preserving the required network span- as recommended by the manufacturer).

REFERENCES

Fruit Color and Shape Grading Using a New Vision System

H. Golnabi*1
1Institute of Water and Energy
Sharif University of Technology
Tehran, Iran

R. Jafari2
2Plasma Physics
Islamic Azad University
Tehran, Iran

A. Haghhighatzadeh2
2Plasma Physics
Islamic Azad University
Tehran, Iran

E. Jali2
2Plasma Physics
Islamic Azad University
Tehran, Iran

Abstract

A CCD based vision sensor system is described in this study. Such a vision system offers both the color images from the object under study and also provides quantitative information concerning the image intensity profile. It is possible to obtain reflection and image data by using single or double light sources for the object illumination. Using the active and passive lighting combination, the reflection data and image profile for different fruits are obtained and presented in this study. To check the potential of the reported vision system for fruit grading and shape analysis some of the related data is described here. The image intensity profile as a function of the x and y coordinates are obtained from fruit surfaces such as banana (plain surface) and apple surface (spherical). The observed intensity profile for the plane surface is uniform in both x and y directions while the spherical surface shows a maximum value at a particular x and y values. In another study the reflected light from fruit surfaces are collected by using a laser light source. Reflected intensity depends on the light wavelength, surface color and its characteristics, which can be helpful in vision analysis. Proposed method can be implemented effectively in the fruit production line for color and shape/size grading purposes.

1. INTRODUCTION

Expected functions from a machine vision system are the capturing of images, analysis of those captured images, recognition of certain objects and features within each image, and the initiation of subsequent actions in order to accept or reject the corresponding objects. Vision system stages are image acquisition, pre-processing of image, segmentation, feature extraction, classification, inspection, and finally actuation which is an interaction with the scene under study. Vision system developments in manufacturing can result in improvements in the reliability, in the product quality and enabling technology for a new production process [1]. Operation of a vision system starts with the image acquiring in which image sensing, digitization, and representation of the image data are accomplished. Scene constraint is the main consideration for the machine vision system and the situation of the scene must be recognized by the machine vision designer. Thus according to the required application needs an appropriate machine vision system should be developed. The hardware for this system consists of the light source for the active imaging and required electro-optical and opto-mechanical sub-systems.

Different lighting techniques such as structured lighting can be used for scene illumination. One possible development in the area of vision hardware system is using the laser scanning systems. To acquire 3-D information one can use a bi-dimensional scanning method as described in Ref. [2]. Another possibility is to use x-ray or infrared lighting, which offers some advantages. For example machine vision and infrared lighting technique has been used to track production quantities, detect defects, and screening problems. The infrared spotlight is employed to detect and inspect glass-bottle flaws [3]. X-ray technology machine vision system has been proposed for the inspection of the air-bag inflators. High quality machine vision system should integrate an image acquisition subsystem that produces a high resolution pictures (512x512 pixels). Such devices offer significant advantage to acquire precise information about relative positions of several objects within a scene in a single glance. Design of industrial machine vision...

*H. Golnabi: Tel.: +98(21) 6616-4652; Fax: +98 (21) 6601-2983; E-mail: golnabi@sharif.edu
systems for automated manufacturing and the role of laser lights were presented in our recent reports [4-7].

2. VISION SYSTEM DESIGN

Block diagram of a typical vision system is shown in Fig.1. As can be seen it includes a light source illuminating on the object under study, an image CCD camera, the interface card for the camera, and a PC. The output of camera is connected via an interface board to a PC and signal processing and camera control is accomplished by using related software [1]. Several tasks such as image acquisition, processing, segmentation, and pattern recognition are conceivable using machine vision system. In recent years a new field of applications in sensing area has opened up and using such fibers a variety of optical sensors have been reported as physical and chemical sensors. The goal has been here to implement such fiber design in the construction of vision sensors/systems.

As can be seen in Fig.1, it includes two light sources illuminating on the object under study via a dual fiber design and direct illumination, a CCD camera, the interface card for the camera, a PIN photodetector, a digital multimeter, and a PC. In the double fiber design one fiber is used for the illumination of the object and the second one is used for the transmission of the reflected light from the object to the digital multimeter. A silicon photodiode (BPX 65 Centronic) is used for light detection and conversion of optical signal to an electric one. The electric output signal of this detector is connected to a digital voltmeter by a coaxial cable. A digital multimeter (SANWA Electronic Instrument Co. LTD, PC 5000 DMM) is used for the output voltage reading and data processing (±0.1mV precision). This device equips with an optical isolated interface port at the meter back for data communication with a PC. The software (PC Link plus) allows one to log measuring data into a PC through RS232 port. The operation of this software is possible by using any operation system such as windows 98, NT4.0/2000/ME/XP versions.

The output of CCD camera is connected via an interface board to a PC and signal processing and camera control is accomplished by using related software. The CCD camera is a miniature camera (34×34 mm size, model TBC-312C, TÉVICOM). It is a camera with alternate lens configuration (flat pin-hole, Con type pin-hole lens) 1/4 inch color interlines transfer CCD, built in microphone, extremely low power consumption, and high quality image. Other specification of this camera is a scanning system (2-1 interlace 625 Lines/50 Fields, 25 Frames), effective pixels 510(H) × 491(V), resolution of 380 TV Lines, illumination minimum of 0.8 LUX, S/N ratio >48 dB. Video output 1 V p-p 75 ohm, VBS, with auto white balance, background compensation and internal synchronization. It has an electronic shutter 1/60-1/100000, lens regular, flat pin-hole, Con pin-hole lens. Power source requirement is 12 VDC and 200 mA with a power consumption of 4.5 W. The used lens option is Con-pin-hole lens (F2.0:f3.6) type that its selection depends on the application requirements [8].
The CCD camera operates with the DVD maker USB2.0 hardware and PVR PLUS software (KWORLD) that operates under all the Microsoft Windows operating systems. The DVD maker USB2.0 hardware requires a Pentium-IV 1.6GHz or higher speed type, 256MB RAM of system memory or above, a free USB port, Graphic card, CD-ROM drive, and needs 1GB of free HD space. After the drivers are installed, the next step is to install the PVR PLUS software. The software functions are the TVR: this function turns the computer PC into a full-featured video recorder. MPEG Encoder: MPEG encoder converts AVI files into MPEG format. Video Editor: makes editing MPEG video files easy. Function Burns DVD: burn DVD creates VCD, SVCD and DVD that can be played on a PC or DVD player. The output of the CCD camera is a PGM file type, which is often hard to be processed by the regular programs. In this experiment arrangement is made to convert the portable graphical (PGM) files to MATLAB files, which can be easily manipulated by the MATLAB or EXCEL program. A typical matrix dimension of 240 rows by 320 columns is provided in which the pixel data of the camera are stored.

It is possible to use the coherent and non-coherent light sources for the object illumination. A with lamp operating under 20-25 V dc voltage is used in this experiment. By variation of the supply voltage, it is possible to change the brightness of the light source. Typically a voltage of about 25 V is used in the results reported here. However it is important to have a similar source light intensity for illuminating all the objects under study. Otherwise the obtained results can not be compared together without the required normalization. One of the important design parameter in developing a machine vision system is the most effective type of lighting, which depends on the light source and the related optics. Now system integrators can select from a number of illumination technology such as fiber optics, tungsten lamps, fluorescent, and light emitting diodes in a number of possible configurations. These configurations in clude ring light, spotlight, backlight, and diffuse light. Light sources are able to enhance imaging contrast. Proper light source helps discern imaging features for machine-vision systems. The reflectance light measurement is used for the simultaneous monitoring of both reflected and transmitted lights. Even though the reflectance image could also be monitored by a similar CCD camera, but in this case we used a photodiode arrangement to record the reflected intensity of the light from the objects under study. More information concerning design and applications of vision system are given in Refs. [9-12].

3. Experimental Results

For data acquisition, the MATLAB program is executed in which different subprograms and parameters concerning the data acquisition, and the characteristics of the image are entered as input. By running this program the picture taking begins by the camera while the recording time depends on the camera speed and the number of frames captured. The goal here has been to get simultaneous reflective and transmittance information from the object under study. In the first experiments only light source 1 is used in Fig.1. The aim here is to find a transmitted image signal for the test samples illuminated by the white light source. Furthermore it is desired to have stored image data for further analysis. The combination of the live image and the stored one can results in a true picture of the object under study, which is the ultimate goal of any ideal vision system. Considering this point in this study the live image and its corresponding stored data are recorded. Since there are many images and data only typical examples are described in this report. In order to cut the effect of the room light in the reflection measurements a black cover is prepared for the experimental set up. In this case the room light is also off and the corresponding image signal is due to the dark noise generated by the CCD camera.

A low noise CCD usually offers a lower background noise signal. A white light source and the photo-electronic detections the ratio of the signal to ratio (S/N) is the defining factor for the signal detection. In order to increase this ratio the signal level should be increased and the equivalent noise signal must be reduced as much as possible. However, in increasing the signal level one must avoid the saturation of the photo-detectors. In this experiment care was taken not to saturate the CCD camera photo-sensors. For this reason the lower light intensity was used for the picture taking and image construction. In data representation the capture program normalizes the image intensity to one and care was taken to record intensities lower than one to avoid the detector saturation.

3.1 SHAPE GRADING

To check the potentials of the reported vision system for grading and size analysis some of the related preliminary data are presented in this section. Fig. 2 displays an intensity profile from the taken image for the yellow banana using a white light source. The intensity profile as a function of x is shown in the Fig.2. As can be seen in Fig.2, a uniform intensity is resulted for such a scan since the banana skin is almost uniform in dimension through the scanned distance of about 6 mm. A similar investigation is performed for the y direction and the results are shown in Fig.3. Similar to
the x direction a uniform intensity profile is obtained from the banana skin in the given direction. It was noticed that for the fruits with a planar surface the intensity profile is nearly uniform as indicated in Fig. 2 and Fig. 3.

![Figure 2: Intensity profile for the banana skin in the x-direction](image)

![Figure 3: Intensity profile for the banana skin in the y-direction](image)

Similar studies are performed for the apple with the spherical surface profile. Fig. 4 displays an intensity profile from the taken image for the red apple using a white light source. The intensity profile as a function of x is shown in the
Fig. 4. As can be seen in Fig. 4, the observed intensity shows a maximum value (58 mV) at x value of about zero. From this point as can be seen in Fig. 4, the intensity decreases in an exponential way to reach a value of about 50 mV as shown in Fig. 4. A similar investigation is performed for the y direction and the results are presented in Fig. 5. Similar to the x direction, an intensity profile is maximum of about 57 mV at a y-value of about zero and reach to a value of about 48 mV at y=3.5 mm.

**Figure 4:** Intensity profile for the reflection from apple skin in the x-direction

**Figure 5:** Intensity profile for the reflection from apple skin in the y-direction
3.2 COLOR GRADING

In order to test the potential of the reported vision for color grading some of the related preliminary data are presented. The sample tests for this study are the fruits with different colors. As can be seen, the reflection data is very sensitive to the surface material and color that can be used for color recognition and grading of different fruits and agricultural products. Such study also can provide some information concerning the skin properties of the fruits or other products. However, more details about such applications require more experiments in this field that is one of our future goals for the reported design. Of course, some modification is required in order to fit the reported system for any particular application depending on the application requirements. Such modification can be in the area of light illumination and sensing images.

In the next study, reflection and image data resulting from real-size fruits are presented. Fig. 6 shows the real-time CCD images of different fruits at a similar condition.

Figure 6: CCD images of different fruits. (a) green apple room light off, (b) green apple, (c) yellow plum, and (d) orange nectarine
In order to get some information from reflection in Fig. 7 reflection signals for different fruit samples are shown. As we know the reflected light depends on the light and surface characteristics. In Fig. 7, since the structure and the reflectivity is different from one type to another one; therefore it is hard to contribute the signal difference to only the fruit color surface.

4. CONCLUDING REMARKS

A general-purpose machine vision system with its preliminary applications was described here. Design and operation of the reported vision system is based on the combination of the room light and a source light. By using both the reflection and image intensity profile techniques data for the object under study are obtained and analyzed. Using this system it is possible to obtain the real-time image of the object under study or store the image data in a matrix for the further processing. In real-time analysis it is possible to look at different processes and display images in different formats. In post-processing it is possible to average image signals from different shots, look at the background signal and correct the resulting image signals for the background corrections.

Object identification and classification are one of the most important applications of a vision system. Applying reported vision system can be helpful in the analysis of small objects and monitoring production processes. It also provides on-line inpection possibility and imaging options for fruits and packing in automotive production assembly lines. Sorting of parts, identifying and unloading of parts from pallets are important applications. Sorting and grading of food and other products are an other example of such identification applications. A list of possible applications of the vision system with more details can be found in literature [13-17].
The next generation vision system should implement laser source light, in particular semiconductor lasers, in order to improve the quality of the captured images. The next generation should lead towards machine vision systems with the universal capabilities to provide 3-D description of the scene. In the design and operation of a vision system, image formation and visual processing, computational methods and algorithms, depth information, image representation, and modeling and matching must be considered. On the other hand, systematic consideration is important in the efficiency and performance of the selected machine. Integration possibility, robustness, ease of operation, and adding intelligence into the system in order to make it a smart system are features of an advanced machine vision system.

ACKNOWLEDGEMENTS

This work was supported in part by the Sharif University of Technology research program. The authors gratefully acknowledge the grant devoted to this research.

REFERENCES

Feasibility of an AIDC System within a Manufacturing SME

Sara Zarei*, Simon Hodgson and Farhad Nabhani

School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

Automatic Identification and Data Collection (AIDC) Systems have been widely used within the manufacturing industry to allow information to pass easily and accurately between different stages of the supply chain. AIDC technologies can be categorised as Biometrics or Data Carrier technologies; Biometrics include fingerprint recognition technologies, facial recognition and voice recognition whereas Data Carrier consists of optical, magnetic or electronic storage. AIDC systems utilising Data Carrier technologies have been successfully implemented within many large multinational organisations but research into the implementation of AIDC systems within SMEs is far more limited.

The processes and operations conducted at a manufacturing SME located in the North East of England were critically analysed and explored in order to identify areas where lean manufacturing principles can be implemented and also areas where the use of an AIDC system could be used to improve the efficiency and visibility of the processes throughout manufacture. Process activity charts were produced for each of the processes at the company to include the operations and transfer times for analysis. This provided a systematic approach for the analysis of current processes and for determining the cost benefits based on the identification and reduction of non-value added operations in the process. The areas for improvement could then be identified and solved through specific applications and/or systems of which a cost benefit analysis could be conducted. The feasibility of the implementation of an AIDC system and the recommended modifications to the shop floor layout were then discussed to determine the benefits available through the recommendations made.

1.0 INTRODUCTION

The case study will be conducted at a Small to Medium Enterprise (SME) located in Middlesbrough, in the North East of England that specialise in the manufacture and refurbishment of feedscrews and barrels for the plastic, rubber and food industries and also provide expertise in welding and hardfacing for the vessel and nuclear industries. [1, 2] The company manufacture parts as required through a specific customer order, known as job shop manufacturing, which provides a high level of customer support and interaction.

In the manufacturing industry, the tracking of a product and its status through the processes of a company, from the moment it arrives until the time of despatch, can significantly reduce the lead time, increase the efficiency and visibility, increase the accuracy of production plans and also improve the efficiency of an ERP system. As the constant demand for total visibility and fast and accurate information flow increases, the inefficiencies of traditional systems are becoming more apparent. [3]

This paper aims to educate and assist organisations with the selection process of AIDC technologies within a Small to medium size manufacturing organisation. A process mapping exercise is conducted to investigate areas that may benefit from an automatic identification system and the improvements that could be made. A method of assessing the cost benefits of an RFID system application is also covered as this is considered the major barrier for the implementation of AIDC systems within SMEs. This study aims to establish a methodology that may be used by many organisations in order to investigate areas of improvement within a process that may benefit from the implementation of such a system.
2.0 Automatic Identification and Data Collection Systems

Automatic Identification and Data Capture (AIDC) technologies and systems were initially developed as a response for the requirement to collect and transfer levels of data in an efficient, quick and accurate manner. AIDC technologies have since developed immensely and may now be used for many purposes in a number of industries. AIDC technologies can reduce or even eliminate the occurrence of errors in the collection of data and the manual input or entry of data and has been successfully applied in the retail, logistics and manufacturing industries. AIDC technologies are distinguished in two separate groups; biometrics, or feature extraction, and data carrier technologies as shown in Figure 1 [3 – 5]

![Figure 1 – AIDC Technologies](image)

Biometric or feature extraction technologies are AIDC technologies that measure physical and behavioural characteristics by exploiting object, action or property relations. The most widely used feature extraction technologies include fingerprint recognition, facial recognition and voice recognition but technologies such as iris and retinal scanning are becoming more popular with the increased technological capabilities. Data Carrier technologies encode and decode information through three separate means; optical storage technologies, magnetic storage technologies and electronic storage technologies. [3, 6-7]

2.1 Optical Storage

Linear Barcodes - The most widely used AIDC technology is linear barcodes which are predominantly used in the retail industry and can be found on almost all commercially available products as the technology is easy to implement and has a low installation cost. Linear barcodes have also found applications in manufacturing support systems such as inventory control, tool control, shipping and transportation. The barcodes encode data via width or height modulated lines and spaces that are of a parallel arrangement and data may only be encoded in one direction. Width modulated barcodes encode the data through the variation in the widths of the lines and spaces present, whereas height modulated barcodes encode the data through the variation in the height of the lines. The size of the barcode utilised is dependant on the amount of data that it is required to carry; the more data required, the larger or longer the barcode. [3][4][8]

2D Barcodes – The limitations of linear barcodes, with respect to the amount of data they can carry and their size, have made them inadequate for some applications as they do not have the capability to carry the level of data required in the traceability applications required in industries such as the automotive, aerospace and pharmaceutical industry. Large volumes of data are necessary to be encoded, sometimes in size limiting spaces, in order to keep track of manufactured parts, assembled parts, supplier information and batch and cost information. The requirement to carry this large amount of data and the often limited space available has led to the development of 2D. [3][5][8]

Optical Character Recognition - OCR is the use of a scanning device to identify a printed text through a computer system. The use of certain fonts, such as OCR-A and OCR-B, comprising of numbers 0-9, letters A-Z and a limited number of characters are the means of information transfer. OCR technologies have currently had very little impact on traditional AIDC applications due to their previously high costs and limitations in recognition algorithms. [3][9]
2.2 Magnetic Storage

Magnetic stripes are the second most widely used AIDC technology as they are often seen in credit/debit cards, railway tickets and identification cards. This technology encodes the data onto a magnetic stripe, which is made from many small magnetic particles bound together in a resin based compound, and can be applied to any non-magnetic surface material. There are three tracks contained within the magnetic stripes onto which the data may be encoded; the characteristics of these tracks, their functional usage and their position have a pre-defined standard. The density of the bits on each track dictates the amount of information that can be encoded onto it, also taking into account the format of the data. [3][4]

2.3 Electronic Storage

Smart Cards - A portable device which can be used to automatically identify items, transaction and events is known as a smart card. The smart card must have a storage memory, a chip operating system memory, a central processing unit in the form of an integrated circuit chip and a read/write interface in order to function correctly. The data required is stored in the chip operating system memory and then may be accessed for various processing applications. [3]

Radio Frequency Identification (RFID) - Radio Frequency Identification (RFID) is an AIDC technology that is used to identify people or objects carrying encoded microchips through the use of radio waves. RFID technologies were initially developed by the British military in the 1940s in order to be able to distinguish their own aircraft from the enemy aircraft which was performed by equipping the aircraft with devices to send out coded signals in order to identify them as their own. RFID technologies have since found applications in many everyday devices such as remote controls for gates or garage doors, anti theft devices in vehicles, security systems and the new European Union passports. [3][10-11]

3.0 Methodology

In order to determine and sustain areas of improvement within a company, the processes and operations currently in use must be understood and analysed in depth. Process mapping is a visual approach that is used to observe, understand and question each step of the process while analysing any potential areas for improvement. Each step of the process is identified, described in terms of how they relate to one another and all influencing factors are identified. These influencing factors include inputs into the process, outputs of the process, people involved in the process, time taken to complete the process, the quality of the parts produced and the working conditions involved. [12 – 13] There are many different approaches that may be used to conduct a process mapping exercise; however, they all have the same major features. All process mapping approaches aim to identify the different types of activity that take place during the process and illustrate the flow of materials, people and information through the process. [12]

A process chart, or process activity map, can be used to show the sequence of operations or activities involved in the process from beginning to end with the duration of each activity included. A process activity map analyses each operation of the process as well as any delays, transportation time and inspection stages throughout with the aim of distinguishing between different steps in the process and categorising them as an operation, transportation, inspection, delay or storage activity. Using a process activity map, it is also possible to critically analyse and distinguish between the value adding and non-value adding operations for a more detailed evaluation of the process. A value adding operation is an activity that directly adds value to the end part or process as it is something that the customer is willing to pay for, whereas a non-value adding operation is any activity that does not directly add value to the end part or process as the customer will not pay extra for this operation in particular. Non-value adding operations should be attempted to be eliminated from the process, however, they can sometimes be an operation which must be performed for the process to flow but does not directly add value to the process. Such operations cannot be fully eliminated from the process as they are required for the successful completion of the process. Once an accurate process map is complete, an analysis can take place and any areas for improvement can be identified and suggested in an attempt to eliminate any bottlenecks in the system or idle times which would affect the efficiency and productivity. [12]

Due to the fact that the SME studied manufacture on a specialist and precision level specifically to the order of a customer, known as job shop manufacturing, they do not follow a specific sequence of processes for every job. As each job is different, the process mapping exercise for the company became very complex. In order to conduct the process mapping exercise, an example job of each type of product manufactured or refurbished at the company was examined to give an indication of the typical stages followed throughout each process. Upon analysing each stage of
the process for an example of every product manufactured; it became apparent that the flow of materials, people and communications across the shop floor was extremely erratic and disorganised. In order to investigate the flow across the shop floor at a more critical level, a plan of the shop floor was gained and the movements of all stages followed were mapped out. This method allowed a visible interpretation of the transportation steps throughout the company prior to the analysis of the processes themselves.

Once the current state process mapping exercise was completed accurately, the current state of each process was analysed in order to distinguish any bottlenecks and/or wasteful activities that may be eliminated or reduced through the application of an Automatic Identification and Data Collection (AIDC) system. The duration of each stage of the process was also established along with the type of activity documented which allowed for a critical analysis prior to comparison between the current state and modified state of the process. The various steps in the current state process mapping exercise were analysed by following each stage of the process as it took place through direct observation, collecting the data directly from the shop floor and discussing the process with workforce involved. The layout of the shop floor was also identified as a major problem when observing the operations and stages of each process. The positioning of machinery, stores and work bays did not seem to appropriately facilitate the manufacturing process for any of the jobs followed which caused a large amount of wasted time and resources in transportation manoeuvres.

The potential areas where improvements could be made were highlighted by identifying and eliminating the non value adding operations such as manual input of data and manual creation of vast amounts of paperwork. The transportation times were also attempted to be reduced through the alteration of the shop floor layout by improving the visibility, communication and positioning of machinery and work bays. The implementation of a suitable AIDC system was investigated as a method of eliminating the extensive level of paperwork and the transportation of this paperwork from one stage to another throughout the process. The main requirements of the AIDC system were to increase process efficiency, be applicable to the metallic surfaces of the products, utilise handheld portable devices for use on the shop floor and also be user friendly with a minimum amount of training required. The AIDC system aims to replace a various number of manual steps that are currently in place at the company and also increase the visibility and communications between each stage of the process in order to reduce the lead time of a product. The implementation of an AIDC system within the company has the ability to reduce the amount of wasted time throughout the process, reduce the quantity of paperwork and physical inputs required and reduce the level of inventory required whilst also increasing the communication and visibility of parts across the shop floor with the aid of a modified layout. However, in order to establish the actual benefits of the implementation of an AIDC system and the alteration of the shop floor layout, a cost benefit analysis must be performed to establish the savings that could be achieved with such an investment.

### 4.0 RESULTS

#### 4.1 PROCESS MAPPING

The layout of the shop floor at the SME was identified as a problem due to the poor visibility and communication between the stages and also the lengthy transfer times. Due to the fact that the company is split into two separate workshops and each individual stage requires the part to be frequently moved from one to the other, the transfer times throughout the process were high. The design of the shop floor was therefore modified in an attempt to facilitate the processes as best as possible without compromising one for another. The layout was modified so as each stage of the manufacturing process follows the previous in a U-shaped manner across the factory floor from beginning to end; however, even after modifications were made, not all of the process could follow an exact U-shaped production line. The processes were re-mapped for the modified shop floor layout in order to analyse the future flow of materials and information. As the company manufacture parts to a specific customer order, the process stages followed for each job are different, making it impossible to have an exact U-shaped production line for every process. The modified shop floor plan was designed to be the best possible compromise for each of the jobs followed in order to improve the visibility, communication and efficiency throughout the process. Figure 2 shows the current shop floor plan with the flow of processes for one job and Figure 3 shows the modified shop floor layout with the flow of processes for the same job.
The current state process activity map was analysed and each individual operation was considered alongside the benefits that may be felt through the implementation of an AIDC system and the modification of the shop floor layout. The processes were then simplified to become more efficient and reduce any unnecessary stages. The duration of each operation was also critically considered for the areas in which increased visibility and communications through the use of an AIDC system were identified and improved. The reduction in the transfer times between the stages through the modified shop floor layout was also considered in the process activity charts and was adjusted accordingly. The modified state process activity maps could then be created with any unnecessary non-value adding activities eliminated alongside any avoidable delays and lengthy transfer times. Figure 4 shows the differences between the current and modified process maps with the elimination of any delays and lengthy transfer times due to increased communication and visibility through the implementation of an AIDC system and the time reduction in physical transfer time between stages due to the modified shop floor layout.
4.2 COST BENEFIT ANALYSIS

Once the process mapping exercise was conducted for the current and visualised state of the manufacturing process for an example of each product, a cost saving analysis can be carried out to assess the potential savings to be made. In order to analyse the savings as accurately as possible the figures for inventory levels, safety stock and sales figures were gained from the company.

The level of safety stock could be significantly reduced through the implementation of an AIDC system within the company. [14] Using the information and data collected from the company, the estimated savings per year from the reduction of inventory levels through the implementation of such a system are shown in equation (1)

\[
\text{Savings} = (\text{Total Inventory Value} \times \text{Carrying Cost} \%) \times (\text{Safety Stock} \%) \times (50\%) \\
= (£130,000) \times (10\%) \times (50\%)
\]

\[
= £6,500
\]

The cost of manual counting and tracking times can be reduced by 80% through the implementation of an AIDC system. [14] The manual tracking of the jobs across the shop floor was completed on a weekly basis and took 5 hours in total to complete, including the manual input of the results; this also proved insufficient as real time job tracking is required. The estimated savings per year for the elimination of manual tracking is shown in equation (2)

\[
\text{Savings} = (\text{No. of counts}) \times (\text{Time taken per count}) \times (\text{Hourly labour cost}) \times (80\%) \\
= (52) \times (5) \times (10) \times (80\%)
\]

\[
= £2,080
\]

Improvements in process efficiency through the implementation of an AIDC system can also be calculated through the process mapping exercise. The amount of hours saved for each process was gained through the comparison of the modified process activity charts and the current process activity charts. Table 1 illustrates the total annual savings that can be seen through the implementation of a suitable AIDC system at the company. The savings for each process are calculated using equation (3).
Savings = Process duration x Parts per process x Hourly labour cost

\begin{align*}
\text{Table 1 – Cost Benefits for Each Process} \\
\begin{array}{|l|c|c|c|}
\hline
& \text{Hours saved per} & \text{Cost savings per} & \text{Cost savings per year} \\
& \text{process} & \text{process (£)} & \text{ (£)} \\
\hline
\text{New Screw} & 8 & (8) x (1) x (10) = 80 & (80) x (300) = 24,000 \\
\text{New Barrel} & 3 & (3) x (1) x (10) = 30 & (30) x (150) = 4,500 \\
\text{New Feed Liner} & 6.25 & (6.25) x (1) x (10) = 62.5 & (60) x (80) = 5,000 \\
\text{Tie Bar} & 7.75 & (7.75) x (1) x (10) = 77.5 & (77.5) x (50) = 3,875 \\
\text{Screw Refurbishment} & 5.5 & (5.5) x (1) x (10) = 55 & (55) x (350) = 19,250 \\
\text{Barrel Refurbishment} & 4 & (4) x (1) x (10) = 40 & (40) x (50) = 2,000 \\
\hline
\text{TOTAL} & & & \text{£58,625} \\
\end{array}
\end{align*}

Table 2 illustrates the total annual savings that can be considered for this case study through the implementation of a suitable AIDC system for inventory, tracking and process efficiency.

\begin{align*}
\text{Table 2 – Total Annual Savings} \\
\begin{array}{|l|c|}
\hline
\text{Savings} & \text{Amount (£)} \\
\hline
\text{Inventory Savings} & \text{£6,500} \\
\text{Manual Tracking Savings} & \text{£2,080} \\
\text{Process Efficiency Savings} & \text{£58,625} \\
\text{Total Annual Savings} & \text{£67,205} \\
\hline
\end{array}
\end{align*}

4.3 **AIDC System Design**

The level of data required for each part at the company would mean that the linear barcode used would have to be a minimum of 1.5cm high and 9.5cm wide in order to contain it all. [15] They are also easily damaged through scratches and dirt, and therefore were deemed unsuitable for application on to the metal surfaces of the products at the company.

2D barcodes applied to the product in the form of Direct Part Marking (DPM), by permanently engraving the code onto the part, were examined. Data Matrix is found to be the most appropriate method of DPM for use as job tracking as the information encoded may be read even when it has suffered up to 40% damage and would also required a much smaller space than linear barcodes, at about 1cm high and 1cm wide. [15] The quality and durability of the mark is affected by the material into which it is engraved and the environments to which it is exposed. The harsh manufacturing environments of the products at the company would mean that the mark could easily be damaged through the turning, milling and grinding stages of the process.

Externally applied passive RFID tags were unsuitable for use within the SME as the presence of the tag externally would compromise the performance of the process and the functionally of the part itself. Upon a comparison of the AIDC technologies, glass type tags were found to be the most suitable as they could be embedded into the parts with minimal machining and are small enough to be discreet and not interfere with the manufacturing or performance of the part. [16-18] The glass tags come in a variety of sizes, ranging from 12mm to 32mm in length which makes them versatile for parts of all sizes and allows for minimum machining for the application of the tag. An RFID tag would have to be embedded into the body of the product at the company to protect from damage and allow the product to function properly. The metal surfaces of the parts could cause a problem for RFID technologies, as the metallic surfaces can interfere with the signals used to read the tag; however, it was found that the use of an adhesive resin to position the tag within the part would provide protection for the tag and also a barrier layer between the tag and the metal surface which would allow for the data to be read without any interference.

The main restraint for any AIDC system at this company would be the temperature extremities as the screw flights are heated to temperatures of a maximum of 200°C during the plasma hardfacing process and so the tag would be required to be positioned in a different place for this part. If the technology applied is required to be present on the product throughout its lifecycle in the event of refurbishment being required then it can be subjected...
to temperatures ranging between 110°C and 320°C in use. However, there are areas on each product that are not subjected to these extreme temperatures, such as the drive of the screw, which would only experience temperatures up to a maximum of 90°C. The tags would need to be applied to the area of the product that is exposed to the least heat. The AIDC technology utilised must be able to withstand certain levels of oil, grease and dirt along with levels of damage that could occur throughout the manufacturing operations that the product would be exposed to. The application of an embedded glass tag with an adhesive resin for protection has the ability to withstand such conditions. The fact that the tag is embedded within the part would avoid levels of damage through machining and the adhesive resin used with the glass tag would provide a waterproof and protective coating that would not allow grease, dirt or liquids to reach it. [17-18]

5.0 CONCLUSIONS

The process activity maps highlighted the areas in which the implementation of an AIDC system can add value and the aim of eliminating non-value added operations allowed improvement opportunities to become apparent. The process mapping exercise also provided information that aided the calculation of the cost benefit analysis for an AIDC system through the number of hours saved per process due to improvements. However, due to the fact that the SME pursue job shop manufacturing and make each product specifically to a customer order, it is difficult to predict that each job in the future will have the same timescales or follow the same sequence of operations. The layout of the shop floor at the company was modified in an attempt to facilitate the manufacturing and refurbishments processes as best as possible without compromising one for another. The modified shop floor plan was therefore designed to be the best possible compromise for each of the jobs followed in order to improve the visibility, communication and efficiency throughout the process. The costs associated with the movement of machine and disturbance of production will have to be outweighed against the benefits of the improved layout.

The accurate traceability of products throughout the manufacturing process can improve the efficiency of a process, the accuracy of the information available and improve the lead time of the product. The traceability of parts may be improved through the implementation of an RFID system with components that are capable of withstanding the conditions of the part. Passive LF glass type RFID tags were identified as a viable solution as they may be embedded into the metallic parts produced and can withstand all conditions felt through the products lifecycle. The cost benefit analysis for the SME through the implementation of an RFID system and the modification of the shop floor plan was found to provide a saving of £67,205 annually. Significant cost savings are found through the implementation of an RFID system based on the reduction of safety stock, the elimination of manual job tracking and the reduction of the manual input and written data throughout the process. The modification of the shop layout better facilitated the use of an RFID system as well as improved the efficiency of the process by providing increased visibility and a large reduction in transportation times.

ACKNOWLEDGEMENTS

With thanks to all employees at the SME used in this case study for their input and guidance throughout the course of this project.

REFERENCES

1. SME Company Brochure
2. SME Company Website
7. Understanding AIDC in Manufacturing Production and Control – Identifying Areas and Processes Where the Technologies Can Be Used. 50-58. 34


Design of a Cobot with Three Omni-Wheels

Z. M. Bi
Northern Ireland Technology Centre
Queen’s University Belfast
Belfast, BT9 5HN, UK

Lihui Wang
Centre for Intelligent Automation
University of Skövde, PO Box 408
541 28 Skövde, Sweden

ABSTRACT

Collaborative robots (Cobots) have been proposed to guide and assist human operators to move heavy objects in a given trajectory. Most of the existing cobots use steering wheels; typical drawbacks of using steering wheels include (i) the difficulty to follow a trajectory with a curvature larger than that of the base platform, (ii) the difficulty to mount encoders on steering wheels due to self-spinning of the wheels, and (iii) the difficulty to quarantine dynamic control performance since it is purely kinematic control. In this paper, a new cobot with the omni-wheels has been proposed, and its design model has been developed, and a simulation has been conducted to validate this control performance.

1. INTRODUCTION

The collaborative robot (Cobot) is invented by Peshkin and Colgate [1] at Northwestern University. A cobot is a robot designed to assist human beings as a guide or assistor in a constrained motion. In contrast to traditional mobile robots, the cobots provide the guidance rather than moving power. Therefore, they are passive in the sense that the power for fulfilling a moving task is provided by operators. However, the wheels of a cobot are motorized to confine their moving directions when an external force is applied to drive them from planned trajectories. Unlike the robots that perform sophisticated tasks in the restricted areas isolated from operators, the cobots have been designed to work with human operators in a shared workspace [2]; moreover, cobots allow human operators use their muscle power more safely and effectively [3,4].

Most of the cobots are developed for two primary purposes: (i) to assist operators in moving heavy objects under some physical constraints, and (ii) to assist patients or disabilities in navigation or rehabilitation. For purpose (i), Peshkin et al. [5] have proposed a variety of cobots for materials handling. Some cobots for assembly become commercially available from Stanley [6]. Hodgson and Emrich [7] and Townsend et al. [8] have developed and patented a prototype cobot for the surgeons to move C-arm x-ray machines, where the surgeon prefers not to hand over control of a procedure to an autonomous robot. For purpose (ii), Ulrich and Borestein [9] have developed a guide-cane to navigate visually impaired patients. Nejatbakhsh et al. [10] have developed a passive walker for rehabilitation.

Existing cobots mostly use the steered wheels, and their structures are similar to mobile robots with the steered wheels. However, the rolls of the wheels are free in a cobot rather than being controlled in a mobile robot. Therefore, the control mechanisms are significantly different from that of a mobile robot. A few of researchers have studied the controls of the cobots. Peshkin and his collaborators have established the most extensive control models for tricycle cobots. In their works, trajectories are planned in terms of the parameterized arc length, and the rolling speeds of steered wheels have to be detected to decide control parameters [1,3,11-14]. Most of the cobots have three or more wheels; but one or two wheels can also be applied. For example, Ulrich and Borestein [9] have proposed a cobot with two wheels to assist walking for visually impaired patients. It is our observation that a cobot with the steered wheels has the following drawbacks: (i) the difficulty to follow a trajectory with a curvature larger than that of the base platform, (ii) the difficulty to mount encoders on steering wheels because of its self-spinning, and (iii) the difficulty to quarantine the dynamic control performance since it is purely kinematic control.

* Corresponding author: Tel.: +46 (0500) 44-8519; Fax: +46 (0500) 44-8598; E-mail: lihui.wang@his.se
To overcome the aforementioned drawbacks, we propose to substitute the steered wheels with omni-wheels. Consequentially, a significant difference is that the omni-wheels will be actually controlled by braking torque instead of steering, and dynamic behaviours and corresponding models have to be developed to implement its control. No research has been found on the development of a cobot with the omni-wheels. In fact, very limited research has been conducted to develop the mobile robots using omni-wheels. Nejatbakhsh et al. [10] is the only research group who has been working on the design of the passive walker with four omni-wheels; however, their primary works have been focused on the system identification and the modification of the dynamic characteristics based on a patient’s walking ability. In this paper, a cobot with three omni-wheels is proposed, the kinematic and control dynamic models are established, and the feasibility of its control is evaluated via simulation.

2. CONCEPTUAL SYSTEM DESIGN

The conceptual design of the cobot is shown in Figure 1, the cobot platform is supported by three symmetrically-distributed omni-wheels. Each omni-wheel has a motor allowing the torque applied along its primary rotate axis. At the centre of the cobot platform, the reflectors of a motion tracker are mounted to detect the position and orientation of the cobot platform. An operator will apply the moving force and torque through the operator’s handler, and a 6 DOF (degree of freedom) sensor is used to detect the force and torque of the operator.

![Figure 1: Conceptual design of a cobot with three omni-wheels](image1)

![Figure 2: Cobot parametric model](image2)

The parametric model of the cobot is shown in Figure 2. The motion of the cobot can be described with respect to a world coordinate system \( \{O_w-X_wY_w\} \), and the orientation of the cobot platform is described by three parameters \( \{x_c, y_c, \theta_c\} \) at the origin \( O_c \). In the same way, the centre position of an omni-wheel is described by \( \{x_i, y_i, \theta_i\} \). A local coordinate system \( \{O_c-X_cY_c\} \) is defined on the mobile platform. The locations of the omni-wheels are defined by the parameters \( L_i \) and \( \alpha_i \). \( m_p \) and \( m_i \) are the masses of the mobile platform and an individual omni-wheel, respectively, and \( I_p \) and \( I_i \) are their inertias.

3. KINEMATIC MODEL

The kinematic model describes the relationship between the motions of the omni-wheels and that of the cobot platform. Forward kinematics is concerned in this section; forward kinematics maps a prescribed motion trajectory of the cobot platform into the motions of the omni-wheels.

3.1. THE MOTION OF THE PLATFORM

As shown in Figure 2, the motion of the platform is denoted as \((r_c, \theta_c)\), where \( r_c \) is the origin \( O_c \) of the cobot, and \( \theta_c \) is the z-axis rotation of the cobot platform under the world coordinate system \( \{O_w-X_wY_w\} \). A motion along the prescribed trajectory is described as,

\[
x_c = x_c(t), \quad y_c = y_c(t), \quad \theta_c = \theta_c(t) \quad \text{or} \quad x_c = x_c(s), \quad y_c = y_c(s), \quad \theta_c = \theta_c(s)
\]

where \( t \) is the time variable and \( s \) is the scalar variable representing the arc length along the prescribed trajectory. \( r_c = [x_c, y_c]^T \) is the position of the reference origin of the cobot platform, and \( \theta_c \) is the orientation of the cobot platform.

The velocities and accelerations of the cobot platform with respect to the time \( t \) are as follows,
Design of a Cobot with Three Omni-Wheels

\[ \dot{x}_c = \frac{dx}{ds}, \quad \dot{y}_c = \frac{dy}{ds}, \quad \dot{\theta}_c = \frac{d\theta}{ds} \]

(2)

\[ \ddot{x}_c = \dot{k}_{c_1} + \dot{k}_{c_2} \dot{x}, \quad \ddot{y}_c = \dot{k}_{c_1} + \dot{k}_{c_2} \dot{y}, \quad \ddot{\theta}_c = \dot{k}_{c_1} + \dot{k}_{c_2} \dot{\theta} \]

(3)

where \( k_{c_1} = \frac{d^2x}{ds^2} x^2, k_{c_2} = \frac{d^2x}{ds^2} \), \( k_{c_1} = \frac{d^2y}{ds^2} y^2, k_{c_2} = \frac{d^2y}{ds^2} \), and \( \dot{x}_c, \dot{y}_c, \dot{\theta}_c \) are calculated from eq.(8). Similarly, the acceleration of \( x, y, \theta \) and \( x, y, \theta \) are obtained.

Note that \( \dot{x}_c, \dot{y}_c, \dot{\theta}_c \) depend only on the prescribed trajectory and they can be obtained directly from eq. (1) if the trajectory is known.

\[ \dot{x}_c, \dot{y}_c, \dot{\theta}_c \] will affect the linear velocity \( \dot{x} \) and acceleration \( \dot{y} \) of the motion along the trajectory.

3.2. Derivation of the Motion with Respect to an Arbitrary Parameter

It is common that an explicit description of a trajectory requires a curve parameter other than time or arc length. Therefore, the conversion between different descriptions has to be determined. Assume an arbitrary curve parameter \( q \) has to be used to describe a trajectory explicitly, i.e.

\[ x_s = x_s(q), \quad y_s = y_s(q), \quad \theta_s = \theta_s(q) \]

(4)

Deviating with respect to time \( t \) gets,

\[ \dot{x}_s = \frac{dx_s}{dq} \frac{dq}{dt}, \quad \dot{y}_s = \frac{dy_s}{dq} \frac{dq}{dt}, \quad \dot{\theta}_s = \frac{d\theta_s}{dq} \frac{dq}{dt} \]

(5)

\[ \ddot{x}_s = \frac{d^2x_s}{dq^2} \left( \frac{dq}{dt} \right)^2 + \frac{d^2x_s}{dq^2} \frac{dq}{dt}, \quad \ddot{y}_s = \frac{d^2y_s}{dq^2} \left( \frac{dq}{dt} \right)^2 + \frac{d^2y_s}{dq^2} \frac{dq}{dt}, \quad \ddot{\theta}_s = \frac{d^2\theta_s}{dq^2} \left( \frac{dq}{dt} \right)^2 + \frac{d^2\theta_s}{dq^2} \frac{dq}{dt} \]

However, one has,

\[ \frac{ds}{dt} = \sqrt{\left( \frac{dx_s}{dt} \right)^2 + \left( \frac{dy_s}{dt} \right)^2} = \frac{d^2s}{d\theta_s^2} = \frac{\dot{x}_s + \dot{y}_s}{\sqrt{\dot{x}_s^2 + \dot{y}_s^2}} \]

(6)

Substituting eq. (5) into eq. (6) gives \( \frac{d\theta_s}{dt} = \frac{d^2q}{dt^2} \).

3.3. The Motions of the Omni-Wheel Centres

The position \( r_i \) of the centre of an omni-wheel can be determined from the known posture of \( O_i \) as,

\[ r_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} x_i + L_i \cos(\theta_i + \alpha_i) \\ y_i + L_i \sin(\theta_i + \alpha_i) \end{bmatrix} \quad (i = 1, 2, 3) \]

(7)

Taking the derivation of eq. (7) with the time \( t \) gives the velocity of \( r_i \) as

\[ \frac{dr_i}{dt} = \begin{bmatrix} 1 & 0 & -L_i \sin(\theta_i + \alpha_i) \\ 0 & 1 & L_i \cos(\theta_i + \alpha_i) \end{bmatrix} \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \]

(8)

where \( \dot{r}_i = (\dot{x}_i, \dot{y}_i) \) and \( \dot{\theta}_i \) are calculated from eq.(8). Similarly, the acceleration of \( r_i \) is obtained as

\[ \frac{d^2r_i}{dt^2} = \begin{bmatrix} \ddot{x}_i \\ \ddot{y}_i \end{bmatrix} = \begin{bmatrix} k_{x_i} + k_{y_i} \ddot{x} \\ k_{x_i} + k_{y_i} \ddot{y} \end{bmatrix} \]

(9)

where \( k_{x_i} = \left( \frac{d^2x_i}{ds^2} - L_i \frac{d\theta_i}{ds^2} \cos(\theta_i + \alpha_i) \right) \ddot{x} - L_i \frac{d\theta_i}{ds^2} \sin(\theta_i + \alpha_i), \)

\( k_{y_i} = \left( \frac{d^2y_i}{ds^2} + L_i \frac{d\theta_i}{ds^2} \sin(\theta_i + \alpha_i) \right) \ddot{y} - L_i \frac{d\theta_i}{ds^2} \cos(\theta_i + \alpha_i), \)

\( k_{x_i} = \frac{d^2x_i}{ds^2} - L_i \frac{d\theta_i}{ds^2} \cos(\theta_i + \alpha_i), \)

\( k_{y_i} = \frac{d^2y_i}{ds^2} + L_i \frac{d\theta_i}{ds^2} \sin(\theta_i + \alpha_i) \)

Eqs. (7-9) describe the motion of an omni-wheel with respect to the world coordinate system \( \{O_w-X_wY_w\} \). However, it is convenient to take it into consideration in a local coordinate system of the omni-wheel. The conversion between two motion descriptions is as follows

\[
\begin{bmatrix}
\dot{x}_i \\
\dot{y}_i
\end{bmatrix} = \begin{bmatrix}
ca_i&sa_i \\
-sa_i&ca_i
\end{bmatrix}
\begin{bmatrix}
x_i \\
y_i
\end{bmatrix},
\begin{bmatrix}
\dot{x}_i \\
\dot{y}_i
\end{bmatrix} = \begin{bmatrix}
ca_i&sa_i \\
-sa_i&ca_i
\end{bmatrix}
\begin{bmatrix}
x_i \\
y_i
\end{bmatrix},
\begin{bmatrix}
\ddot{x}_i \\
\ddot{y}_i
\end{bmatrix} = \begin{bmatrix}
ca_i&sa_i \\
-sa_i&ca_i
\end{bmatrix}
\begin{bmatrix}
\ddot{x}_i \\
\ddot{y}_i
\end{bmatrix} = \begin{bmatrix}
k_{x,i1} + k_{x,i2} \ddot{x}_i \\
k_{y,i1} + k_{y,i2} \ddot{y}_i
\end{bmatrix}
\]

(10)

where \( k_{x,i1} = k_{x,i1}ca_i + k_{y,i1}sa_i, \ k_{x,i2} = k_{x,i2}ca_i + k_{y,i2}sa_i, \ k_{y,i1} = -k_{x,i1}sa_i + k_{y,i1}ca_i, \ k_{y,i2} = -k_{x,i2}sa_i + k_{y,i2}ca_i \).

Figure 3: The force analysis of an omni-wheel
Figure 4: The force analysis of the cobot platform

4. DYNAMIC CONTROL MODEL

The torques on the primary rotation axes of the omni-wheels is controlled to confine the cobot platform in the prescribed trajectory. The actual acceleration along the trajectory depends on the external forces applied by an operator and applied torques on the omni-wheels.

4.1. FORCE ANALYSIS OF AN OMNI-WHEEL

The force balance in an omni-wheel is illustrated in Figure 3. The omni-wheel is connected to the cobot platform by \( X' \)-axis of its primary rotation. The reacting force and torque from the cobot platform to the omni-wheel are \( F(F_x,F_y,F_z) \) and \( \tau \). Besides, the omni-wheel has an external normal reacting and friction force \( N_i \) and \( (f_x,f_y) \), respectively, from the ground.

Taking into consideration of the force balance of the omni-wheel gives,

\[
\begin{bmatrix}
F_x \\
F_y \\
F_z
\end{bmatrix} + \begin{bmatrix}
f_x \\
f_y \\
N_i - mg
\end{bmatrix} = \begin{bmatrix}
m_r \ddot{x}_i \\
m_r \ddot{y}_i \\
0
\end{bmatrix}
\]

(11)

where \( F\) \( (F_x,F_y,F_z) \) is the reacting force from the platform to the omni-wheel, \( f_x,f_y \) is the friction force from the ground, \( N_i \) is the normal reacting force from the ground, \( (\ddot{x}_i,\ddot{y}_i) \) is the translational acceleration of the omni-wheel defined in eq. (10), and \( m_r \) is the mass of the omni-wheel.

Taking into consideration of the torque balance gives,

\[
\begin{bmatrix}
\tau \ \\
0
\end{bmatrix} + \begin{bmatrix}
f_x R_x \\
f_y R_y
\end{bmatrix} = \begin{bmatrix}
I_{x} \ddot{\theta}_x \\
I_{y} \ddot{\theta}_y
\end{bmatrix}
\]

(12)

where \( \tau \) is the torque applied to \( X' \)-axis of the omni-wheel, \( I_x \) and \( I_y \) are initial moments of the omni-wheel with respect to \( X' \) and \( Y' \) axes, respectively.
\[ \ddot{\theta}_y = \ddot{\gamma}_y / R_y \quad \text{and} \quad \ddot{\theta}_x = \ddot{\gamma}_x / R_x \] are the angular accelerations of the omni-wheel with respect to the \( X' \) and \( Y' \) axes, respectively. \( R_y \) and \( R_x \) are the radius of the omni-wheel and its rolls.

From eqs. (11-12), \( f_{x_i}, F_{\gamma_i} \) can directly be calculated as,
\[ f_{x_i} = I_y \ddot{x}_y / R_x^2 \]
\[ F_{\gamma_i} = \left( m_x - I_y / R_y^2 \right) \ddot{\gamma}_x \] (13)

Substituting eq. (10) into eq. (13) gives,
\[ f_{x_i} = k_{f_{x_{i1}}} + k_{f_{x_{i2}}} \ddot{s}_{x_i} \]
\[ F_{\gamma_i} = k_{F_{\gamma_{i1}}} + k_{F_{\gamma_{i2}}} \ddot{s}_{\gamma_i} \] (14)

where \( k_{f_{x_{i1}}} = I_y \ddot{x}_y / R_x^2 \), \( k_{f_{x_{i2}}} = I_y \ddot{x}_y / R_x^2 \), \( k_{f_{\gamma_{i1}}} = (m_x - I_y / R_y^2) \ddot{\gamma}_x \), \( k_{f_{\gamma_{i2}}} = (m_x - I_y / R_y^2) \ddot{\gamma}_x \).

\( f_{x_i}, F_{\gamma_i} \) can be expressed in terms of the unknown variable \( \tau_i \) as,
\[ F_{\gamma_i} = m_y \ddot{\gamma}_y + \left( \tau_i - I_y \ddot{\gamma}_y / R_y \right) / R_y \]
\[ f_{x_i} = \left( -\tau_i + I_y \ddot{\gamma}_y / R_y \right) / R_y \] (15)

Substituting eq. (10) into eq. (15) gives,
\[ f_{x_i} = k_{f_{x_{i1}}} + k_{f_{x_{i2}}} \ddot{s}_{x_i} + k_{f_{\gamma_{i1}}} \ddot{s}_{\gamma_i} + k_{F_{\gamma_{i1}}} \ddot{s}_{\gamma_i} \]
\[ F_{\gamma_i} = k_{F_{\gamma_{i1}}} + k_{F_{\gamma_{i2}}} \ddot{s}_{\gamma_i} + k_{F_{\gamma_{i1}}} \ddot{s}_{\gamma_i} + k_{F_{\gamma_{i2}}} \ddot{s}_{\gamma_i} \] (16)

where \( k_{f_{x_{i1}}} = I_y \ddot{x}_y / R_x^2 \), \( k_{f_{x_{i2}}} = I_y \ddot{x}_y / R_x^2 \), \( k_{f_{\gamma_{i1}}} = (m_x - I_y / R_y^2) \ddot{\gamma}_x \), \( k_{f_{\gamma_{i2}}} = (m_x - I_y / R_y^2) \ddot{\gamma}_x \), \( k_{F_{\gamma_{i1}}} = 1 / R_y \), \( k_{F_{\gamma_{i2}}} = 1 / R_y \).

Note that the following constraint applies to prevent the slip:
\[ \sqrt{f_{x_i}^2 + F_{\gamma_i}^2} \leq \mu_N \] (17)

where the normal reacting forces from the ground to the omni-wheels can be determined from the following equations,
\[ \sum_{i=1}^{3} N_{x_i} + F_{ec} x = 0 \]
\[ \sum_{i=1}^{3} N_{y_i} = (m_x + 3m_e) g + F_{ec} \]

\[ \text{4.2. FORCE ANALYSIS OF THE PLATFORM} \]

As shown in Figure 4, the cobot has an external force \( F_e \) and torque \( \tau_e \) applied by a human operator. The force and torque are uncontrollable. A force sensor has to be applied to measure their values and directions in real-time. When the cobot platform is isolated from the omni-wheels, the reacting forces \( -F_{x_i}, -F_{y_i}, -F_{\gamma_i} \) and torque \( -\tau_i \) from the omni-wheels are considered to be external.

Taking into consideration of the force balance on the plane of \( X_wY_w \) with respect to the local coordinate system on the platform gives,
\[ F_e - \sum_{i=1}^{3} \begin{bmatrix} c\alpha_i & -s\alpha_i \\ s\alpha_i & c\alpha_i \end{bmatrix} F_{x_i} = m_x \begin{bmatrix} c\theta_e & -s\theta_e \\ s\theta_e & c\theta_e \end{bmatrix} \tau_e \] (18)

In eqs. (14) and (16), the forces on the omni-wheels are obtained in the local coordinate systems of the omni-wheels. They have to be transformed into the same coordinate system, i.e., the local coordinate system on the cobot platform.
Substituting eqs. (6), (14) and (16) into eq. (18) gives,

\[
F_{x,x} = -\sum_{i=1}^{3} \left[ ca_i \left( k_{x,i} + k_{x,i} \dot{s} \right) - sa_i \left( k_{y,i} + k_{y,i} \dot{s} + k_{z,i} \tau \right) \right] = m_p \left( k_{x,1} \theta_x - k_{y,1} \dot{\theta}_y + k_{z,1} \dot{\theta}_z \right) + \left( k_{x,2} \theta_x - k_{y,2} \dot{\theta}_y + k_{z,2} \dot{\theta}_z \right) \dot{s}
\]

\[
F_{y,y} = -\sum_{i=1}^{3} sa_i \left( k_{x,i} + k_{x,i} \dot{s} \right) + ca_i \left( k_{y,i} + k_{y,i} \dot{s} + k_{z,i} \tau \right) = m_p \left( k_{x,1} \theta_x + k_{y,1} \dot{\theta}_y + k_{z,1} \dot{\theta}_z \right) + \left( k_{x,2} \theta_x + k_{y,2} \dot{\theta}_y + k_{z,2} \dot{\theta}_z \right) \dot{s}
\]

\[
(19)
\]

The above equation can be changed into the following form,

\[
U_{11} \tau_1 + U_{12} \tau_2 + U_{13} \tau_3 + U_{14} \dot{s} = U_{15}
\]

\[
U_{21} \tau_1 + U_{22} \tau_2 + U_{23} \tau_3 + U_{24} \dot{s} = U_{25}
\]

where \(U_{ij}\) is the co-efficient determined from eq. (19).

Taking into consideration of the torque balance with respect to \(Z_w\) gives,

\[
\tau_e + F_e \times r_e - \sum_{i=1}^{3} \left[ ca_i \left( k_{x,i} + k_{x,i} \dot{s} \right) - sa_i \left( k_{y,i} + k_{y,i} \dot{s} + k_{z,i} \tau \right) \right] F_{i,x} \times r_e = I_p \dot{\theta}_i
\]

\[
(21)
\]

where \(I_p\) is the inertial moment of the platform with respect to \(\{O_c-X_c-Y_c-Z_c\}\), \(r_e\) is vector from the reference origin to the location of the force is applied and detected, and \(\tau_e\) is external torque with respect to \(Z_w\) applied by a human operator.

Substituting the eqs. (14) and (16) into the eq. (21) gives,

\[
U_{31} \tau_1 + U_{32} \tau_2 + U_{33} \tau_3 + U_{34} \dot{s} = U_{35}
\]

\[
(22)
\]

where \(U_{3i} = k_{j_i} \dot{L_i}, U_{34} = -I_p k_{g,2} - \sum_{i=1}^{3} k_{j_i} \dot{L_i}, U_{35} = -I_p k_{g,1} - F_e \times r_e - \sum_{i=1}^{3} k_{j_i} \dot{L_i}
\]

Combining eqs. (20) and (22) makes a set of three linear equations. When \(\tau_i\) \((i=1,2,3)\) are dealt with design variables, they can be solved and represented by \(\ddot{s}\) as,

\[
\tau_i = k_{x,i} \ddot{s} + k_{y,i} \ddot{s} \quad i = 1,2,3
\]

\[
(23)
\]

where \(k_{x,i}, k_{y,i}\) are the coefficients determined \(U_{ij}\) in eqs. (20) and (22).

4.3. Constraints to maintain the ground contact

When the torques applied on the axes of the omni-wheels, they will change the distributions of the normal reacting forces from the ground. These normal reacting forces depend on the external force by the human operators and the torques applied on the omni-wheels. The model of determining these forces has been developed but not included in this paper. These forces have to be positive so that appropriate contacts can be maintained between the cobot and the ground.

5. Case study

As shown in Figure 9, the output of the control is the motion acceleration required to follow the planned trajectory, and the inputs of the control include (i) the prescribed trajectory that can be revised based on the real-time feedback, (ii) the detected position and orientation of the cobot platform, (iii) the detected translational and rotational velocities of the cobot platform, and (iv) the detected force and torque applied by a human operator.

In the conceptual design, the following input parameters are assumed,

(i) the omni-wheels are distributed symmetrically as \(L_i = 0.3\ m, \alpha_1 = -\pi/6, \alpha_2 = -\pi/2, \alpha_3 = 7\pi/6;\)
(ii) the platform has a mass $m_p=600$ kg, and $I_{pc} = m_p L^2/4$;
(iii) the location of the force sensor is $r_e = (0, L/2, 0)$;
(iv) the parameters of an omni-wheel are $R_o=0.02$ m, $m_o=0.5$ kg, $I_{ao}=m_o R_o^2/4$;
(v) the parameters of a roll on an omni-wheel are $R_e=0.005$ m, $m_e=0.05$ kg, $I_{ae}=m_e R_e^2/4$;
(vi) the required trajectory is defined in term of the trajectory parameter $q$ as
   \[ x = 0.125q + 0.8 \cos(q) - 1.0, \quad y = 0.8 \sin(q), \quad \theta = 0.1q; \]
(vii) the external forces are assumed as,
   \[ \angle(F, Y) = 0.25 \pi (1 - 2 \times \text{rand}()), \]
   \[ F_x = [100 + 50(1 - 2 \times \text{rand}())] \cos(\angle(F, Y)), \]
   \[ F_y = [100 + 50(1 - 2 \times \text{rand}())] \sin(\angle(F, Y)), \]
   \[ F_c = [-100 - 50(1 - 2 \times \text{rand}())], \]
   \[ \tau_c = 0.05(1 - 2 \times \text{rand}()). \]

where $\text{rand}()$ represents a randy number between (0.0, 1.0) and $\angle(F, Y)$ is the angle between the project of $F_e$ on the plane $X_Y$ and $Y$. The gravity acceleration and the friction coefficient are assumed to be 9.81 $kg/m^2$ and 0.1, respectively. The simulation results from the developed control model are demonstrated in Figures 5-8.
The graphic simulation model has been developed for the validation of the control model. The results have shown that it is feasible to control a cobot with the omni-wheels to follow a specified trajectory by applying controlled torques on the primary axes of the omni-wheels. The applied torques will be transformed into the reacting friction forces from the ground to the omni-wheels, and the reacting friction forces will be combined with the external forces and torque applied by an operator to generate a motion acceleration along the given trajectory. Figure 8 shows that the applied torques are not always to decelerate the wheel rotation; either positive or negative torque can be required. In particular, when the trajectory includes a rotation of the cobot platform, the torques applied on the omni-wheels at two sides are opposite. Graphic simulation has been developed to validate the controlled motion of a cobot platform, and Figure 10 depicts a screen shot of the cobot motion in simulation.

6. CONCLUSIONS

In this paper, a new cobot with the omni-wheels is proposed, its kinematic and dynamic models are developed, and the simulation is conducted to illustrate its feasibility. The cobot can be used as a vehicle to move a bulky device along a specified trajectory. A unique feature of the new cobot is its redundant control; the number of the control variables is more than the DOF of the cobot motion. Note that the existing cobots are purely passive and operated manually, and some fully automated active robots have their disadvantages in terms of the safety and cost issues. The most significant advantage of the new hybrid cobot is that the degrees of manual or machine operations of the system can be adjusted in the control model. Therefore, the modes of the manual operation, the full automation, and in-between can be selected and switched based on operation requirements.

REFERENCES

Surface Monitoring by Color Reflectance Measurements

H. Golnabi \(^1,2\)  
\(^1\)Institute of Water and Energy  
Sharif University of Technology  
Tehran, Iran  
\(^2\)Plasma Physics Research  
Islamic Azad University  
Tehran, Iran  

A. Haghighatzadeh\(^2\)  
\(^2\)Plasma Physics Research  
Islamic Azad University  
Tehran, Iran  

S. Dabagh\(^2\)  
\(^2\)Plasma Physics Research  
Islamic Azad University  
Tehran, Iran

**ABSTRACT**

Diffusion reflection measurements from different plane surfaces using a laser diode light are reported. A double-fiber arrangement is used in which one fiber transmits the laser light to the sample surface and the second one transmits the reflected light to the photodetector for the signal measurement. The novelty of reported design is that it compactness of the fiber probe, its simplicity, and it is possible to use the source and the detector at the same side. The intensity measurements for the transparent dielectric, matte plastic surfaces and colored papers are presented in this report. For the thick color plastics, the yellow color has the highest reflectance and the black one shows the lowest. For the thin colored papers the yellow paint shows a maximum diffuse reflectance, whereas, the gray has the lowest diffuse reflectance. Our results in general agree well with the theoretical predictions and also with the other experimental results of diffuse reflection at a near source wavelength. Reflection information reported by this method can be useful in analysis of laser material processing and imaging processes considering the role of reflected light. It can be also useful in design of displacement sensor for object tracking and recognition, vision systems in automotive industry or any automated industry.

1. INTRODUCTION

When a ray of light strikes the surface of an object, it may be absorbed, transmitted, or reflected. Considerable attention has been paid to reflectance study of various surfaces. Some studies have concentrated on the experimental determination of surface reflectance, and at the same time, many theories have been developed for surface reflection modeling [1]. Two limiting cases have been considered at the early stages: the light reflected in a particular direction (perfect specular), or the light reflected in all directions of a hemisphere above the surface uniformly (Lambertian). These two ideal cases for a surface are never attained in practice, and it is advisable to consider both effects to describe the reflection properties of a surface. The distribution of the reflected light in each case described depends on the light direction, wavelength, polarization, and details of the microstructure of the surface layer. For the symmetric surfaces, the reflection is not altered by rotating a surface path about its normal, while in the case of asymmetric surfaces, there is a preferred direction (diffraction in grating). Many surface layers are symmetric and this results certain degree of simplification of the analysis.

In most experiments light from lamps or laser sources have been used to measure surface properties. The use of laser light in the study of metallic surfaces has been reported by Tanner [2]. Optical methods including specular and diffuse reflection measurements have been given. Considering the importance of both specular and diffuse reflectance, in this article we report the relative reflection measurements from surfaces made from different materials, which can provide new and useful information in this respect.

\(^*\)H. Golnabi: Tel.: +98(21) 6616-4652; Fax: +98 (21) 6601-2983; E-mail: golnabi@sharif.edu
2. THEORY OF REFLECTION

A reflection theory, which considers the surface radiance as a function of both incident and reflected beam, is given by Berthold et al. [3]. In that model, a function denoted by $f_r$ indicates the information about how bright a surface will appear, when viewed from a given direction. The bidirectional reflection distribution function (BRDF), defined in that report, shows the ratio of reflected radiance in a direction toward a portion of the surface. A number of theoretical approaches have been developed to explain the scattering characteristics of rough surfaces [4]. Sinusoidal, sawtooth, rectangular, and random profiles have been used in such studies. Another model considers the particle theory in which a powder sample is treated as a collection of uniformly sized rough surfaced spherical particles. The effect of roughness on the specular reflection has led to the development of many models. A surface is modeled as a collection of specular vertices in which the surface normals are assumed to be normally distributed. This is known as the Torrance-Sparrow model [5] that is widely used in the computational vision system and graphics.

It has been shown that geometrical optics is applicable when the surface irregularities are much larger than the wavelength of the incident light. As reported in such cases the Torrance-Sparrow model approximates the physical optics model developed by Beckman and Spizzichino [7]. A number of alternative physical-optics approaches to the problem have been described in the literature [8]. All the above models ignore the effect of roughness on the diffuse reflection, which is the major drawback of both theories. For surface layers it is possible that light reflected from two different physical paths. Some light is reflected at the interface, and the second occurs for the light that crosses the interface. Tominaga and Wandell [9] reported a standard surface reflection model and illuminant estimation for inhomogeneous materials by using two independent mechanisms. The result of that study can be used to describe the light reflected from inhomogeneous materials, measured in different viewing geometries, to estimate the relative spectral power distribution of the ambient light.

Scattering from metallic and dielectric rough surfaces is discussed by Sallier and Maystre [10], by using the integral theory of metallic and dielectric gratings. Wolff [11, 12] has reported diffuse-reflectance model for smooth dielectric surfaces. Light reflectance from randomly oriented convex particles with rough surfaces has been reported. They considered large particles in comparison with the wavelength and calculated the differential scattering cross section. In another study absolute diffuse reflection from relative reflectance measurements has been reported by Lindberg [13]. In this article, we present some of the results given by Ginnken et al. [14], which is useful to our discussion. The total diffuse radiance corrected for masking and shadowing effects is given by integration in Ref. [15]. The total radiance according to Ref. [14] can be the sum of the specular and diffuse reflections

$$L_r = C[gL_{rs} + (1 - g)L_{rd}] ,$$

where $L_{rs}$ shows the specular and $L_{rd}$ the diffusion reflection. In Eq. (1) parameter $g$ indicates the balance between the diffuse and specular reflections, changes between 0 and 1. The parameter $C$ accounts for the intensity normalization factor.

The reflection of transmitted light from a surface based on the geometrical optic can be obtained easily. In this case, the image of the transmitted light on the entrance plane (image plane) of the receiving fiber is computed. Three distinct cases can be considered for such analysis. Let us consider the beam cross second on the target plane to be circular in the $xy$-plane, with the $z$-axis showing the direction of the propagation of the beam. Consider the horizontal and $y$-axis as vertical axis in the object and image planes. If $r_1$ and $r_2$ represent the transmission (T-fiber) and receiving fibers (R-fiber), respectively, then the overlap of two circles determines the amount of light received by receiving fiber. First case when there is no crossing of two circles, no specular light is received by the R-fiber. When diameter $AB$ is tangent or inside the larger circle, the light is received. The amount received light depends on the amount of the crossing area between these two circles. Using geometrical relations, one can compute $\Delta S$ for two described cases. The light intensity on the surface is given by

$$I = \frac{P_o}{\Delta S}.$$
Where \( P_0 \) is the incident power transmitted by the T-fiber. The power, \( P \), collected by the receiver fiber can be written as

\[
P = \rho I \Delta S,
\]

(3)

where \( \rho \) shows the overall reflection coefficient for the reflecting surface. The maximum value of the \( \Delta S \) can be equal to the area of R-fiber cross section. For a similar double-fiber such as the one used here the core radius for two fibers are equal. The intensity of light entered in T-fiber varies by the axial distance between the fibers and the target plane. Theoretical and experimental result indicate that by increasing distance from zero, the collected power is increased, at a particular distance is at maximum, and further increase causes a decrease in the collected power. For our case the distance at which the signal is maximum is used for all the reflection measurements.

3. EXPERIMENTAL ARRANGEMENT

The instrumental arrangement for this study is presented in Fig. 1. The light source is a 5-mW laser diode operating at 650 nm wavelength. A silicon photodiode (BPX 65 Centronic) is used for light detection and conversion of optical signal to an electric one. The photodiode is reverse biased to \(-18\) V by using a battery pack and operates in the photoconductive mode. The electric output signal of this detector is connected to a digital voltmeter (0.1 mV precision) by a coaxial cable. For optimum detection, a flexible adjusting mechanism is required, so the reflecting surface is mounted on a transnational stage, which provides a fine and smooth movement of the surface in axial direction. This assembly provides a simple and precise way to collect all the reflected light even by using a small area photodetector. To collect data the output signal of the photodetector is connected to a digital multimeter (DMM) as shown in Fig. 1. The output of this DMM (Sanwa, PC-5000) can be interfaced to a PC.

![Figure 1: Experiment set up including a laser diode, a double-fiber, a photodetector, a digital voltmeter and a PC](image)

As can be seen in Fig. 1, a double fiber is used in this experiment in which in one end two fibers are separated and they are close to each other at the other end. The double end is mounted in a proper holder close to the surface under study. The separated fiber ends are connected to the laser diode source and detector accordingly. The double fiber includes similar plastic step-index fibers. The overall diameter of each fiber is about 2 mm, and the core diameter is about 860 \( \mu \)m. The numerical aperture for such fiber is about 0.41. The double fiber is cut from the purchased spool and the fiber ends are simply cleaved and polished. However, a better treatment of the fiber ends can improve the coupling efficiency of the fiber to source, to surface under study, and finally to photodetector. Since the cross section of the fiber is large enough, therefore, the laser light is directly coupled to the fiber. Such direct coupling is very easy and does not require any lens and thus the coupling efficiency is higher for such a big-core fiber.
The laser power is about 3 mW and the transmitted power to the surface is about 0.3 mW as measured by a laser power meter. The double side of the fibers is fixed firmly in a proper holder in a distance of about 10 mm from the surface under study.

Two groups of samples are used for this experiment and each sample could be easily fixed at its place for measurements. In sample placement care was taken to place the plane sample test in perpendicular position to the optical axis (fiber axis). For the better adjustment, a plane mirror is used in place of the sample and optimum situation is obtained. Our samples include: the plastic materials, and thin and thick color papers with the similar thickness in each category group.

4. EXPERIMENTAL RESULTS

Using the experimental arrangement shown in Fig. 1 the diffusion reflection signal is measured and results are described in this section. Each measurement is repeated at least 5 times and average value was recorded for the related sample is quoted here. Diffuse reflectance measurements from different surfaces are reported in Ref. [15]. In that experiments a He-Ne laser has been used for measurements. As concluded in [15], dependence of the reflected light upon the light characteristics and details of the surface structure leads to a complicated problem of interaction of light from surfaces. The mixed reflectance considering both the diffuse and specular reflectance from a surface must be considered in any surface modeling. Experimentally, measurements of both specular and diffuse reflections are required in order to test a model or to describe a surface. In metallic surfaces usually the specular reflection is the predominant, so one expects that the diffuse reflectance to be very smaller than specular reflectance.

In the first study the light reflected off the different plastic surfaces is measured and the output signals are shown in Fig. 2 and 3. Both thick and thin color plastics are studied in this investigation. Fig. 2 shows the reflection signal measurements for the similar matte plastics with different colors.

![Figure 2: Reflection measurement for the matte plastics with different colors](image)

Here as can be seen in Fig. 2, the white color matte plastic has the highest reflectance (292 mV); while the blue color shows the minimum diffuse reflection (146 mV). In a similar study the relative reflectance for transparent dielectric materials such as plastics, Kapton, and Mylar sheets were examined. Among these materials with different colors the clear blue sheet has the lowest diffuse reflection, while the Mylar sheet shows the highest diffuse reflection.
We have measured the relative diffuse reflectance for some clear dielectric materials. These materials include: a piece of black rubber (2 mm thickness), a piece of printed circuit board dielectric material (1.5 mm thickness), a piece of dark Plexiglas (3 mm thickness), a clear Plexiglas plate (3 mm thick), and white matte plastic (1 mm thickness). Among tested materials, black rubber shows the lowest diffuse reflectance signal, while the white plastic material shows the highest diffuse reflectance signal (figure not shown here).

For the transparent plastic sheet, the reflection results for some different materials are shown in Fig. 3. In this experiment, colorless, red, dark blue, light blue, and green plastic materials are tested. As can be seen in this comparison, the colorless sheet shows the highest diffuse reflection (292 mV), while the green one has the lowest reflection signal of about 132 mV. In any experiment, the reproducibility of the obtained results is a very important factor. To check this point for this experiment as shown in Fig. 2, two series of the experimental results are compared with each other. Looking at the results, it is noted that there is a very good agreement between the result of series 1 and series 2 as plotted in Fig. 2. Such procedure is repeated for all the experimental results reported in this study as can be seen in different figures representing the results. The difference in the obtained results is less than 1% of the full scale value, which is a good reproducibility for such measurements. Another important factor in such measurements is the stability of the experimental results. Operation and time stability of the measuring instrument is checked and shows a very good stability for a time period of about 10 minutes.

In another study, we considered different plastic materials and the results for thick materials are presented in Fig. 4. The reflection signals are measured for such different thick color materials with a thickness of about 0.5 mm. Among different colors as indicated in graph bar of Fig. 4, the yellow color has the highest diffuse reflectance (303 mV); while the black color shows the least diffuse reflection of about 113 mV.
Figure 4: Reflection signal measurements for thick plastics with different colors

Figure 5: Reflection signals for the thin color papers
In the next study we considered paper materials and the results for thick, thin, and glossy papers are presented in this study. Fig. 5 shows the reflection signals measured for thin color papers. The relative diffuse reflectance for the colored thin papers with a thickness of about 0.5 mm is measured. Among different colors the yellow one has the highest diffuse reflectance (300 mV); while the gray color shows the least diffuse reflection of about 208 mV.

A similar study was accomplished for the colored papers and the results are shown in Fig.6. For different colors the yellow one has the highest diffuse reflection (304 mV), while the blue color has the minimum diffuse reflection (115 mV) at this wavelength. For comparison the reflection for the standard A4 paper is also shown in Fig. 5 and Fig. 6.

![Figure 6: Reflection signals measured for thick color papers](image)

![Figure 7: Reflection signal measurements for the glossy papers with different colors](image)
In order to compare the results of this study for the color thin papers (Fig.5), and for thick color papers (Fig.6) the results have been normalized to the output of the A4 white paper as indicated in both figures. The reflectance signals of the similar glossy papers are also investigated in this study (see Fig.7). The result of this study indicates that glossy cream paper has the highest diffuse reflection (327 mV) while the blue color shows the minimum diffuse reflection (134 mV). It must be pointed out that both color and material types control the reflection signals as described in this study.

5. CONCLUSIONS

In many applications such as in vision system, laser scanners, laser imagining, laser Doppler measurements, and laser material processing the reflection of the light from the surface of illumination is of great interest. This article described the diffuse reflection from different plane surfaces. Results of this work together with the information provided for the specular reflectance can be helpful in the field of for example computer graphics and vision systems, which are based on image formation from reflected light from an object surface. Such reflection information can be also useful in design of non-contact displacement sensor for object tracking and recognition in automotive industry or other automated industries. Sensing operation can provide information about the position of the part and vision system, in particular, can provide information about the shape of surface under study. Resulting data can be used to extract information such as physical properties of different materials and layers as well.

ACKNOWLEDGEMENTS

This work was supported in part by the Sharif University of Technology research program. The authors gratefully acknowledge the grant devoted to this research.

REFERENCES

Semi-Open Queuing Network Approach for Modeling Autonomous Vehicle Storage and Retrieval System

Banu Y. Ekren¹, Sunderesh S. Heragu¹*, Ananth Krishnamurthy², Charles J. Malmborg³

¹Department of Industrial Engineering
University of Louisville
Louisville, KY, 40292, USA

²Department of Industrial and Systems Engineering
University of Wisconsin-Madison, Madison, WI 53706, USA

³Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute
Troy, NY 12180-3590, USA

ABSTRACT

In this paper, we present an analytical model for autonomous vehicle storage and retrieval system (AVS/RS). The system is modeled as semi-open queuing network (SOQN). An SOQN consists of jobs, pallets and servers. Each job is paired with a pallet and the two visits the set of servers required for processing the job in the specified sequence. In the network of an AVS/RS, storage/retrieval (S/R) transactions are jobs and the autonomous vehicles are pallets. If an S/R transaction requires a vertical movement it uses lift. The lifts and horizontal travel times to/from a storage space are the servers. First, we describe all possible scenarios and their probabilities to derive the general service times. Second, we combine all the service times of the network. Last, we solve the SOQN using an approximate method and obtain the performance measures. The case study is completed for a company that uses AVS/RS in France.

1. INTRODUCTION

Autonomous vehicle storage and retrieval systems (AVS/RS) represent a relatively new technology for automated unit load storage systems. Autonomous vehicles function as storage/retrieval (S/R) devices. Within the storage rack, the key distinction of AVS/R systems relative to traditional crane-based automated storage and retrieval systems (AS/RS) is the movement pattern of the S/R device. The vehicles in an AVS/RS share a fixed number of lifts for vertical movement and follow rectilinear flow patterns for horizontal travel.

It is crucial to design an AVS/RS in such a way that it can efficiently handle the current and future demand requirements while avoiding bottlenecks and having excess capacity. Due to the relative inflexibility of the physical layout and the equipment, it is important to design it right the first time. To be able to evaluate all possible design scenarios, an efficient, valid model that is not time consuming is required. The purpose of this study is to develop an analytical model for AVS/RS.

2. BACKGROUND AND DISCUSSION

An AVS/RS transports a unit load using rail guided automated vehicles (AVs) that follow three-dimensional rectilinear movement patterns. The major system components of the AVS/RS technology are lift mechanisms that are mounted along the periphery of the storage rack to provide vertical movement. The dominant cost component of the AVS/RS technology is the vehicle. Lifts costs are approximately 25% of those for the vehicles.

In most unit-load S/R systems, space-conserving random storage policies are used because of capital cost considerations [1]. Under a random storage policy, the location of a specific load in a storage system is a random variable over time. This policy ensures space occupancy rate maximization by allowing different loads to occupy the same addresses at different times [2].

* Corresponding author: Tel.: +1 (502) 852-2741, Fax: (502) 852-5633; E-mail: s.heragu@louisville.edu
There are various studies that evaluate the performance of AVS/RSs ([3], [4], [5], [6], [7], [8]). Different from those studies, we model the AVS/RS with a SOQN approach. We use an approximate based solution to solve the SOQN. We use ARENA 12.0 a commercial simulation software, to develop the simulation results.

3. SOQN NETWORK MODELING OF AVS/RS

An SOQN consists of jobs, pallets and servers. Each job is paired with a pallet and the two visits the set of servers required for processing the job in the specified sequence. The AVS/RS queuing system can be modeled as a SOQN. In the network of an AVS/RS, storage/retrieval (S/R) transactions are jobs and the autonomous vehicles are pallets. The lifts and horizontal travel times to/from a storage space are the servers and the transactions are customers. Each transaction needs vehicle before it can enter the network. Thus, the total number of transactions in service or waiting for service in the network cannot be larger than the number of vehicles. Figure 1 shows the SOQN model of the AVS/RS. If a vehicle is available, then an arriving transaction enters the network of servers immediately, along with this vehicle. Otherwise, it waits in the transaction queue until a vehicle becomes available. When a transaction exits the system, the vehicle returns to a ‘vehicle’ pool.

![Figure 1: Semi-open queuing network](image)

SOQN assumptions and formulations are given below:

1. The dwell point of a vehicle is the place where the last storage or retrieval transaction is completed.
2. The dwell point of the lift is where the last vertical movement is completed.
3. Available locations are chosen randomly for storage of unit-loads.
4. The number of aisles, columns (bays) and tiers of an actual installation are used to derive the probabilities.
5. The arrival rates for storage and retrieval transactions are independent Poisson processes and equal.
6. The transactions are served by the vehicles on a first come first served (FCFS) rule. The vehicles requiring lifts for vertical movement are also served by FCFS order.

The notation used in the AVS/RS model is given below:

- $A$: number of aisles on each tier
- $T$: number of tiers
- $W$: the width of one storage bay
- $V$: the number of vehicles
- $Y_x$: the distance from the 1st bay to the cross aisle
- $v_L$: the velocity of the lift
- $\lambda_S$: the arrival rate of storage transactions per hour
- $C$: number of bays (columns) per aisle
- $D$: the distance between two aisles
- $H$: the height of one tier
- $L$: the number of lifts
- $T_T$: load or unload transfer time between the lift and the I/O point.
- $v_v$: the velocity of the vehicle
- $T_{LU}$: the time to load/unload to or from the storage
\( \lambda_R \): the arrival rate of retrieval transactions per hour

Because of the agreement with the company, we do not give the specific values of the parameters defined above, except the below parameters:

\[
\begin{align*}
A &= 42 & L &= 7 & C &= 27 \\
T &= 7 & \lambda_S &= 225 \text{ unit-loads} & \lambda_R &= 225 \text{ unit-loads}
\end{align*}
\]

3.1 SCENARIOS FOR MOVEMENT KINEMATICS

To be able to derive the expected service time of a transaction, all probable scenarios are considered. In this section, we define all probable scenarios of transactions and their probabilities.

3.2 SIX TRAVEL SCENARIOS FOR STORAGE TRANSACTIONS

We define six travel scenarios and their probabilities for the storage transactions. \( Sf \) shows that the requested storage location is on the first tier. \( Sn \) means the requested storage location is not on the first tier. \( Vo \) and \( Vf \) mean that the seized vehicle is at the I/O point, and at the first tier but not at the I/O point, respectively. And, \( Vn \) indicates that the seized vehicle is not on the first tier.

**Scenario 1 (VnSn)** - The seized vehicle is not on the first tier and the requested storage position is not on the first tier: Under this condition, the expected travel distance is calculated by considering five movements: (i) Vehicle travels from its current position to the lift (ii) Lift travel from its current tier to the vehicle’s tier (iii) Vehicle travel in the lift to the I/O tier (iv) Loading and lift travels to the destination tier (v) Vehicle travels to the destination storage and unloads the unit-load.

**Scenario 2 (VnSf)** - The seized vehicle is not on the first tier and the requested storage position is on the first tier. Under this condition, the expected travel distance: (i) The vehicle travels from its current position to the lift (ii) Lift travel from the current tier to the vehicle’s tier (iii) The vehicle travels to the I/O (iv) Unit-load is loaded on the vehicle (v) The vehicle travels to the destination storage unloads the unit-load.

**Scenario 3 (VfSn)** - The seized vehicle is on the first tier and the requested storage position is not on the first tier. Under this condition, the expected travel distance: (i) Vehicle travels from its current position to the lift (ii) Lift travels to the first tier (iii) Unit-load is loaded and, lift and vehicle travel to the destination tier (iv) The vehicle travels to the storage rack and unloads the unit-load.

**Scenario 4 (VoSn)** - The seized vehicle is at the I/O point and the requested storage position is not on the first tier. The expected travel distance: (i) Lift travels from its current tier to the first tier (ii) Unit-load is loaded on the vehicle and, the lift and vehicle travel to the destination tier (iii) The vehicle travels to the storage rack and unloads the unit-load.

**Scenario 5 (VfSf)** - The seized vehicle is on the first tier and the requested storage position is on the first tier. The expected travel distance: (i) The vehicle travels from its current position to the lift (ii) Lift travels from its current tier to the first tier (iii) Unit-load is loaded on the vehicle (iv) The vehicle travels to the destination storage unloads the unit-load.

**Scenario 6 (VoSf)** - The seized vehicle is at the I/O point and the requested storage position is on the first tier. (i) Lift travels from its current tier to the first tier (ii) Unit-load is loaded on the vehicle (iii) The vehicle travels to the destination storage unloads the unit-load.

3.3 STORAGE SCENARIOS’ PROBABILITY CALCULATIONS

The occurrence probabilities for the above six scenarios are calculated in Table 1.
Table 1: Probabilities of six storage scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr (S₁)  = \frac{T-1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (S₄) = \frac{\lambda_R}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
</tr>
<tr>
<td>Pr (S₂)  = \frac{T-1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
<td>Pr (S₅) = \frac{1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
<tr>
<td>Pr (S₃)  = \frac{1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (S₆) = \frac{\lambda_R}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
</tbody>
</table>

3.4 SEVEN TRAVEL SCENARIOS FOR RETRIEVAL TRANSACTIONS

The same calculations are also completed for the retrieval transactions. We also describe the first scenario different from the storage transaction scenarios, \( Rf \) and \( Rn \) mean that the retrieval location is (is not) on the first tier.

**Scenario 1 (VnRn):** The seized vehicle’s and requested retrieval positions are not on the same tier and the retrieval location is not on the first tier.

**Scenario 2 (VoRn):** The seized vehicle is at the I/O point and the retrieval location is not on the first tier.

**Scenario 3 (VnRn):** The seized vehicle and the retrieval location are on the same tier other than the first.

**Scenario 4 (VnRf):** The seized vehicle is not on the first tier and the retrieval location is on the first tier.

**Scenario 5 (VfRf):** The seized vehicle and the retrieval location are on the first tier.

**Scenario 6 (VoRf):** The seized vehicle is at the I/O point and the retrieval location is on the first tier.

**Scenario 7 (VfRn):** The seized vehicle is on the first tier and retrieval location is not on the first tier.

3.5 RETRIEVAL SCENARIOS’ PROBABILITY CALCULATIONS

The retrieval transaction scenarios’ probabilities are obtained as Table 2.

Table 2: Probabilities of seven retrieval scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr (R₁)  = \frac{T-2}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (R₄) = \frac{T-1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
<tr>
<td>Pr (R₂)  = \frac{\lambda_R}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (R₅) = \frac{1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
<tr>
<td>Pr (R₃)  = \frac{1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (R₆) = \frac{\lambda_R}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
<tr>
<td>Pr (R₇)  = \frac{1}{T} \times \frac{\lambda_S}{\lambda_S + \lambda_R} \times \frac{T-1}{T}</td>
<td>Pr (R₇) = \frac{\lambda_R}{\lambda_S + \lambda_R} \times \frac{1}{T}</td>
</tr>
</tbody>
</table>

3.6 CUSTOMER COMBINATION

Each scenario of storage and retrieval transaction has different routes and service times. Therefore, each scenario can be seen as a new transaction (customer) type in the queuing system. Thus, we have a multi-product SOQN which has 13 (6+7) types of transactions (customers). When the network servers have general service time distributions, they usually defined by two moments which are the mean and the squared coefficient of variation (scv) of service times. Scv is the ratio of the variance to the mean square.

The multi-class queuing network can be treated as a single-class network using Expressions (1)-(6) (see, [9]). Notations that are used for this technique is given below:

- \( n \) : number of nodes
- \( k \) : number of classes
- \( m_j \) : number of servers at node \( j \)
- \( p_{ij}^r \) : routing probability of class \( r \)
- \( \lambda_0^r \) : external arrival rate of class \( r \)
- \( c_j^2 \) : the scv of the service-time distribution of node \( j \)
\( \lambda_i^r \): arrival rate of class \( r \) to node \( i \)

\( \tau_j^r \): the mean service time of class \( r \) at the \( j \)th node of its route

\( c_{ij}^2 \): the scv of the service-time distribution of class \( r \) at the \( j \)th node of its route

According to that technique first, we obtain the external arrival rates to each node calculated by (1):

\[
\lambda_{0j} = \sum_{r=1}^{k} (b \times \lambda_0^r) \quad \left\{ \begin{array}{l}
 b = 1, \text{ if node } j \text{ is the first node on route class } r \\
 0, \text{ otherwise} \end{array} \right., \forall j = 1,2, \ldots, n
\]  

(1)

The external arrival rate at node \( j \), \( \lambda_{0j} \), is the sum of all arrival rates of classes whose first node is \( j \). And, the flow rate from \( i \) to \( j \) is calculated by (2); the flow from \( i \) out of the network is calculated by (3).

\[
\lambda_{ij} = \sum_{r=1}^{k} (\lambda_i^r \times p_{ij}^r) \quad \forall i, j = 1,2, \ldots, n
\]  

(2)

\[
\lambda_{i0} = \sum_{r=1}^{k} (\lambda_i^r \times p_{i0}^r) \quad \forall i = 1,2, \ldots, n
\]  

(3)

Using equations 2 and 3 the overall routing matrix is obtained by (4).

\[
P_{ij} = \frac{\lambda_{ij}}{\lambda_{0i} + \sum_{l=1}^{n} \lambda_{il}} \quad \forall i, j = 1,2, \ldots, n
\]  

(4)

The combined mean and the SCV of service-time of each node are obtained by (5)-(6).

\[
\tau_j = \frac{\sum_{r=1}^{k} (\tau_j^r \times \lambda_j^r)}{\lambda_j}, \quad \forall j = 1,2, \ldots, n
\]  

(5)

\[
\tau_j (c_{ij}^2 + 1) = \frac{\sum_{r=1}^{k} ((\tau_j^r)^2 \times \lambda_j^r \times (c_{ij}^2 + 1))}{\lambda_j}, \quad \forall j = 1,2, \ldots, n
\]  

(6)

3.7 SERVICE TIME CALCULATIONS

The general view of SOQN of the AVS/RS is illustrated in Figure 2. According to that figure there are two types of queues in the system, one is the vehicle queue and the other is the lift queue. Transactions arriving to the system first enter the vehicle queue, then the lift queue. In the storage transactions, the vehicles travel to the I/O point to pick up the load. In the retrieval transactions, the vehicles travel to the load’s location to retrieve it. For both transactions, the vehicles use a lift if it must travel to a tier other than its. In Figure 2, the server \( VT_1 \) corresponds to the vehicle’s horizontal travel from its current position to the lift’s location. \( LT \) corresponds to the lift’s travel. \( VT_2 \) corresponds to the vehicle’s horizontal travel to complete a storage or retrieval. In other words, it is the travel from vehicle’s current position to the storage/retrieval position after it reaches the destination tier, including load/unload activities. In the real AVS/RS system, the vehicle is seized and not released until it completes the entire storage or retrieval. \( VT_1 \) and \( VT_2 \) are servers used to capture vehicle travel times and thus are not real servers. Therefore, dummy vehicle queues are shown before \( VT_1 \) and \( VT_2 \) in Figure 2.

We also consider a synchronization station after the external queue where transactions and vehicles synchronize. After a vehicle completes storage or retrieval it becomes free and returns to the vehicle pool to be used by the waiting transaction.

Each scenario has its own route and service times at each node. For instance, because the vehicles are at the I/O point in the \( S_4 \) and \( S_6 \) of storage scenarios, the vehicles will not visit \( VT_1 \). In addition, in the \( S_3 \) and \( S_5 \) of retrieval scenarios, the transactions will go to \( VT_2 \) directly after seizing the vehicle.
Consequently, after the derivation of service times of VT₁, LT, and VT₂ of each scenario and combination of them, the mean and the squared coefficient of variation SCV of general service times of all nodes and the routing probabilities P₁-P₇ are obtained as below:

\[ P₁ = 0.464, \quad P₂ = 0.5, \quad P₃ = 0.036, \quad P₄ = 0.659, \quad P₅ = 0.341, \quad P₆ = 0.5, \quad P₇ = 0.5\]

\[ 𝜏₁ = 0.511 \text{ min.}, \quad c₁^2 = 0.115, \quad 𝜏₂ = 0.413 \text{ min.}, \quad c₂^2 = 0.233, \quad 𝜏₃ = 0.756 \text{ min.}, \quad c₃^2 = 0.065.\]

After verifying the VT₁, LT, and VT₂ values using simulation, we complete further analyses and solve the Figure 2 network using Avi-Itzhak’s [10] approximation method:

1. First, we transform the SOQN into a closed queuing network (CQN) except the synchronization station. Then, we identify the synchronization station as the first station where transactions may have to wait for a vehicle.

2. Second, we solve the CQN station using Marie’s technique, and obtain the load-dependent throughput rates. Marie has presented an efficient technique for the approximate analysis of general CQN. The basic idea is to approximate the performance of the general network by the performance of a product-form closed queuing network. It provides satisfactory results for a wide range of queuing networks [11].

3. Third, we replace the synchronization station by a load-dependent exponential server with the service rates obtained in step 2.

4. Finally, we find the mean number of transactions in the external queue and network by considering synchronization station in isolation and by solving the birth and death process (M/M/1) that describes its behavior.

**4. CONCLUSION**

In this paper an analytical model of AVS/RS system is presented. We model the AVS/RS with a SOQN approach. First, the mean and SCV of service times are derived using probabilities of pre-defined scenarios of AVS/RS. Second, the AVS/RS system is modeled as a SOQN by considering a synchronization station in the network. Then, the network is solved using a four stepped iterative method as in Avi-Itzhak (see section 3.3). To obtain the load-dependent throughput rates of the CQN, Marie’s approximation is used [11]. Last, the synchronization station is replaced by a load-dependent exponential server and is solved as a birth and death system. The performance measures are obtained for the mean number of transactions in the external queue and network.

**REFERENCES**


A Method for the Calibration of a 3-D Laser Scanner

Cesare Rossi¹, Vincenzo Niola², Sergio Savino³, Salvatore Strano⁴

¹,²,³,⁴Department of Mechanical Engineering for Energetics (DIME)  
University of Naples “Federico II”  
Naples, via Claudio 21, 80125 ITALY

ABSTRACT

The calibration of a three-dimensional digitizer is a very important issue to take into consideration that good quality, reliability, accuracy and high repeatability are the features which such a good digitizer is expected to have. The aim of this paper is to propose a new method for the calibration of a 3-D laser scanner, mainly for robotic applications. The acquisition system consists of a laser emitter and a webcam with fixed relative positions. In addition, a cylindrical lens is provided with the laser housing so that it is capable to project a plane light. An optical filter was also used in order to segment the laser stripe from the rest of the scene. For the calibration procedure it was used a digital micrometer that move a target with known dimensions. The calibration method is based on the modeling of the geometrical relationship between the 3-D coordinates of the laser stripe on the target and its digital coordinates in the image plane. By this method it is possible to calibrate the intrinsic parameters of the video system, the position of the image plane and the laser plane in a given frame, all in the same time.

1. INTRODUCTION

Calibration is basic requirement in multi-sensor platforms where data needs to represented in a common reference frame for the purpose of analysis and data fusion. On platforms where a camera provides intensity information in the form of an image and a laser supplies depth information in the form of a set of 3-D points, external calibration allows reprojection of the 3-D points from the laser coordinate frame to the 2-D coordinate frame of the image. The process of external calibration is often poorly documented and is almost always notoriously laborious. It usually involves some modification of the scene in the form of markers with features visible by both the camera and the laser. These features are often in the form of edges or corners extracted from the laser data with considerable user intervention. In procedures where the features are edges or corners, the quality of the final estimates is dependent on the accuracy by which features can be localized and so requires the ability to obtain laser data at very high resolution.

The procedure proposed in this document uses no more than the same checkerboard calibration target commonly used for internal calibration of the camera. An interactive GUI allows the user to acquire the images of laser on a target that can move along camera optical axis, by means of a digital micrometer. A robust fitting procedure fits geometrical relationship between the 3-D coordinates of the laser stripe on the target and its digital coordinates in the image plane, based on triangulation principle.

By this method it is possible to calibrate the intrinsic parameters of the video system, the position of the image plane and the laser plane in a given frame, all in the same time.
2. 3-D RECONSTRUCTION: PRINCIPLE OF OPERATION

There are many methods in order to reconstruct the object shape by means of videos or images [1]. In this paper is described a system, that consists of a linear laser emitter and a webcam, and uses triangulation principle applied to a scanning belt on object surface [2].

Camera observes the intersection between laser and object: laser line points in image frame, are the intersections between image plane and optical rays that pass through the intersection points between laser and object. By means of a transformation matrix, it is possible to express the image frame coordinates, in pixel, in a local reference frame. In figure 1 a) is shown a scheme of scanning system: \{W\} is local reference frame, \{I\} is image frame with coordinates system \{u,v\}, and \{L\} is laser frame. \{L2\} is laser plane that contains laser knife and scanning belt on the surface of object, and it coincides with \((x,y)\) plane of laser frame \{L\}, [3].

![Figure 1: reference frames](image)

Starting from the coordinates in pixel \((u,v)\), in image frame, it is possible to write the coordinates of the scanning belt on the object surface, in camera frame by means of equation (1). Camera frame is located in camera focal point figure 1 b).

\[
\begin{bmatrix}
    x_c \\
    y_c \\
    z_c \\
    1
\end{bmatrix} =
\begin{bmatrix}
    \delta_x & 0 & 0 & -\delta_x u_0 \\
    0 & \delta_y & 0 & -\delta_y v_0 \\
    0 & 0 & f & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    u \\
    v \\
    0 \\
    1
\end{bmatrix}
\]

(1)

With:
- \((u_0,v_0)\): image frame coordinates of focal point projection in image plane;
- \((\delta_x, \delta_y)\): physical dimension of sensor pixel along direction u and v;
- \(f\): focal length.

It is possible to write the expression of the optical beam of a generic point in the image frame, that can be identified by means of parameter \(t\).

\[
\begin{align*}
    x_c &= (u-u_0)\delta_x t \\
    y_c &= (v-v_0)\delta_y t \\
    z_c &= ft
\end{align*}
\]

(2)

Laser frame \{L\} is rotated and translated respect to camera frame, the steps and their sequence are:

1. Translation \(\Delta x_{lc}\) along axis \(x_c\);
II. Translation $\Delta y_k$ along axis $y_c$;

III. Translation $\Delta z_k$ along axis $z_c$;

IV. rotation $\phi_k$ around axis $z_c$;

V. rotation $\theta_k$ around axis $y_c$;

VI. rotation $\psi_k$ around axis $x_c$;

Hence in equation (3) the transformation matrix between laser frame and camera frame is:

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\psi_k) & -\sin(\psi_k) & 0 \\ 0 & \sin(\psi_k) & \cos(\psi_k) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta_k) & 0 & \sin(\theta_k) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta_k) & 0 & \cos(\theta_k) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(3)

Laser plane $\{L2\}$ coincides with $(x,y)$ plane of laser frame $\{L\}$, so it contains three points:

$$\{p\}_l = (0,0,0,1)^T; \{q\}_l = (1,0,0,1)^T; \{r\}_l = (0,0,1,1)^T.$$

(4)

In camera frame:

$$\{p\}_c = \begin{bmatrix} x_c - p_x \\ y_c - p_y \\ z_c - p_z \end{bmatrix}; \{q\}_c = \begin{bmatrix} q_x - p_x \\ q_y - p_y \\ q_z - p_z \end{bmatrix}; \{r\}_c = \begin{bmatrix} r_x - p_x \\ r_y - p_y \\ r_z - p_z \end{bmatrix};$$

$$\begin{bmatrix} x_c - p_x \\ y_c - p_y \\ z_c - p_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} - \begin{bmatrix} \Delta x_k \\ 0 \\ 0 \end{bmatrix}.$$

(5)

It is possible to obtain laser plane equation in camera frame, solving equation (6) in $x_c$, $y_c$ and $z_c$:

$$\begin{vmatrix} x_c - p_x & y_c - p_y & z_c - p_z \\ q_x - p_x & q_y - p_y & q_z - p_z \\ r_x - p_x & r_y - p_y & r_z - p_z \end{vmatrix} = 0.$$

(6)

If it is:

$$M_x = \begin{vmatrix} q_x - p_x & q_y - p_y \\ r_x - p_x & r_y - p_y \end{vmatrix}; M_y = \begin{vmatrix} q_x - p_x & q_z - p_z \\ r_x - p_x & r_z - p_z \end{vmatrix}; M_z = \begin{vmatrix} q_y - p_y & q_z - p_z \\ r_y - p_y & r_z - p_z \end{vmatrix};$$

$L_2$ equation in camera frame, is:

$$(x_c - p_x)M_x - (y_c - p_y)M_y + (z_c - p_z)M_z = 0.$$

(7)

It is possible to evaluate coordinates $x_c$, $y_c$, $z_c$ in camera frame, solving system (8) with unknown $t$:

$$\begin{cases} x_c = (u - u_0)\delta_x t \\ y_c = (v - v_0)\delta_y t \\ z_c = ft \\ ((x_c - p_x)M_x - (y_c - p_y)M_y + (z_c - p_z)M_z = 0 \end{cases}.$$
The solution is:

\[ t = \frac{p_xM_x - p_yM_z + p_zM_y}{(u - u_o)\delta_xM_x - (v - v_o)\delta_yM_y + fM_z} \]  

(9)

Equation (9) permits to compute in the camera frame, the points coordinates of the scanning belt on the object surfaces, starting to its image coordinates \((u,v)\). In this way it is possible to do a 3-D objects reconstruction by means of a laser knife.

### 3. DETECTION OF THE LASER PATH ON THE TARGET

A very important step for 3-D reconstruction is image elaboration for the laser path on the target [4, 5, 6], the latter is shown in figure 2.

![Figure 2: laser path image](image)

Image elaboration procedure permits to user to choose some image points of laser line, in order to identify three principal colours (Red, Green, and Blue) of laser line, figure 3.

![Figure 3: laser line RGB representation](image)

With mean values of scanning belt principal colours, it is possible to define a brightness coefficient of the laser line, according to relation (10):

\[ s = \frac{\text{max}(\text{mean}(R),\text{mean}(G),\text{mean}(B)) + \text{min}(\text{mean}(R),\text{mean}(G),\text{mean}(B))}{2} \]  

(10)

By means of relation (11), an intensity analysis is carried out on RGB image.

\[ L(u,v) = \frac{\text{max}(R,G,B) + \text{min}(R,G,B)}{2} \]  

(11)

By equation (11), the matrix that contains the three layers of RGB image, is transformed in a matrix \(L\), that represents image intensity. This matrix represents same initial image, but it gives information only about luminous intensity in each image pixel, and so it is a grayscale expression of initial RGB image, figure 4 a) and b).
A Method for the Calibration of a 3-D Laser Scanner

With relation (12), it is possible to define a logical matrix $I_b$. Matrix $I_b$ indicates pixels of matrix $L$ with a brightness in a range of 15% brightness coefficient $s$.

$$I_b(u, v) = \begin{cases} 1 & \text{if } \frac{0.85}{s} \leq L(u, v) \leq \frac{1.15}{s} \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (12)$$

In figure 4 c), matrix $I_b$ is shown.

In matrix $I_b$, the scanning belt on the object surface (see fig. 5) is represented by means of the pixel value one. The area of the laser path in the image plane depends on the real dimension of laser beam and on external factors such as: reflection phenomena, inclination of object surfaces.

3-D reconstruction procedure is based on triangulation principle and it doesn’t consider the laser beam thickness, so it is necessary to associate a line to the image of scanning belt.

Since the laser path in the image plane is rather horizontal, a geometrical mean is computed on, on the columns of the matrix $I_b$, that is to say: in the opposite direction to wider extension of the laser path in the image. This is shown in equation (13).

$$\Omega := \{u, v\}; \Omega := \{\bar{u}, \bar{v}\} \Rightarrow \bar{u} = \frac{\sum_{u, v \in \Omega} u I_b(u, v)}{\sum_{u, v \in \Omega} I_b(u, v)} \neq \infty, \bar{v} = \nu : \sum_{u, v \in \Omega} I_b(u, v) \neq 0$$  \hspace{1cm} (13)

Then a matrix $h_b$ is also defined as follows:

$$h_b(u, v) = \begin{cases} 1 & \text{if } (u, v) \in \Omega \\ 0 & \text{otherwise} \end{cases} \hspace{1cm} (14)$$

Matrix $h_b$ is hence a logical matrix with same dimension of $I_b$ in which that represents laser image like a line, figure 6. This line is centre line of scanning belt on object surface in image.
The set of transformations (11), (12) and (13) represents the image elaboration, that is necessary to identify a laser line in image. The points of this line are used in 3-D reconstruction procedure.

4. CALIBRATION SYSTEM

It is necessary to identify all model parameters in order to obtain a good 3-D reconstruction. Calibration is a basic procedure for the data analysis [3, 5, 7]. For this reason, it was developed a calibration procedure for a classic laser scanner module.

The laser scanner module is composed by a web-cam with resolution 640x480 pixel and a linear laser. The calibration test rig is realized with a guide on which is fixed the laser scanner module and a digital micrometer with a target, figure 7 a).

A fixed frame $O_{xfyfz}$ has origin in a vertex of the rectangular target with dimension $a=68$ mm and $b=74.5$ mm, like it is shown in figure 7 b). Target movements are indicated with $\Delta z_{bf}$.

In figure 8 some steps of calibration procedure are shown.

Calibration procedure is based on a set of images, taken with the target placed in different positions respect to the laser scanner module. A least square optimization allows to identify all parameters. In figure 9 are shown 20 images used for calibration, of scanning belt, with 20 different $\Delta z_{bf}$. 
A Method for the Calibration of a 3-D Laser Scanner

The aim of calibration procedure is to obtain the parameters that define the relation between laser frame $L_xL_yL_z$ and camera frame $O_xO_yO_z$, figure 1.

In every calibration step, it is possible to define transformation matrix between fixed frame $O_xO_yO_z$ and target frame:

$$\begin{bmatrix} T_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0; 0 & 1 & 0; 0 & 0 & 1 - \Delta \theta; 0 & 0 & 0 \end{bmatrix}$$

(13)

The transformation matrix between fixed frame $O_xO_yO_z$ and camera frame $O_xO_yO_z$ can be obtained analogous to matrix $[T_f]$, and it is a function of six parameters:

$$[T_i] = f(\Delta x, \Delta y, \Delta z, \psi, \theta, \phi) = f(\pi_{cf})$$

(14)

with $\pi_{cf}$ set of parameter of transformation $[T_c]$.

Equation (3) can be written as:

$$[T_i] = g(\Delta x, \Delta y, \Delta z, \psi, \theta, \phi) = f(\pi_{cf})$$

(15)

with $\pi_{cf}$ set of parameter of transformation $[T_c]$.

For each step of the calibration procedure it is possible to evaluate the coordinates of the points of the laser line in the camera frame, according to equations (5) and (13):

$$[\bar{x}, \bar{y}, \bar{z}]^T = f(\pi_{cf}, f) \cdot [\bar{u}, \bar{v}, 0, 1]^T$$

(16)

By means of equation (14), it is possible to write:

$$[\bar{x}, \bar{y}, \bar{z}]^T = [T_i]^{-1} \cdot [\bar{x}, \bar{y}, \bar{z}, 1]^T = [T_i]^{-1} \cdot f(\pi_{cf}, f) \cdot [\bar{u}, \bar{v}, 0, 1]^T = \alpha(\pi_{cf}, \pi_{lc}, f) \cdot [\bar{u}, \bar{v}, 0, 1]^T$$

(17)

The optimization problem can be defined as:

$$\min_{\rho \in R} F(\rho)$$

(18)

$F(\rho)$ is a vectorial function defined as:

$$F(\rho) = \{(\pi_{cf}^{(i)} - \pi_{cf}^{(j)}) - (\Delta \theta^{(i)} - \Delta \theta^{(j)}), \min(x_{cf}^{(i)}), \max(x_{cf}^{(i)}) - \alpha, \max(y_{cf}^{(i)}) - \min(y_{cf}^{(i)}), \max(y_{cf}^{(i)}) - \min(y_{cf}^{(i)}), \min(z_{cf}^{(i)})\}^T$$

(19)

With a least square optimization, the unknown parameters of set $[\pi_{cf}, \pi_{lc}, f]$ are identified.

An interactive GUI was developed to allow the user to acquire images, to execute optimization and to verify the results.
In figure 10, a graphical result of calibration is shown: it is possible to observe the camera frame position, the laser beam (represented by the quadrilateral on the left), and the 3-D reconstructions of scanning beam on target surface, for each image used in the calibration procedure.

Figure 10: calibration results

5. CONCLUSION

A method for calibration of a 3-D laser scanner has been presented. The procedure is absolutely non-invasive since it does not involve any modification of the scene; in fact no markers with features visible by both the camera and the laser, or any other device, are required.

The method essentially uses a laser emitter and a webcam with fixed relative positions. A cylindrical lens and an optical filter are also provided; the first one is in order to project a plane light, while the second one was used to segment the laser stripe from the rest of the scene. The calibration method is based on the modeling of the geometrical relationship between the 3-D coordinates of the laser stripe on the target and its digital coordinates in the image plane. By this method it is possible to calibrate the intrinsic parameters of the video system, the position of the image plane and the laser plane in a given frame, all in the same time.

REFERENCES

A New Real Time Shape Acquisition with a Laser Scanner: First Test Results

Cesare Rossi¹, Vincenzo Niola², Sergio Savino³, Salvatore Strano⁴

¹,²,³,⁴Department of Mechanical Engineering for Energetics (DIME)
University of Naples “Federico II"
Naples, via Claudio 21, 80125 ITALY

ABSTRACT

The first results of a new method for real time shape acquisition with a laser scanner are presented. The new method is essentially based on the use of a laser beam and a web cam. A digital filter parameters identification was studied for the laser line detection in the image. After this a model for the reconstruction in real time of the laser line in the space was developed. The first test rig was just conceived to validate the method, hence no high resolution cameras were adopted. Nevertheless the tests have showed encouraging results.

Tests were made on both plane and non plane surfaces. First of all it was confirmed that it is possible to calibrate the intrinsic parameters of the video system, the position of the image plane and the laser plane in a given frame, all in the same time. Moreover the surface shapes were recognized and recorded with an appreciable accuracy. The tests also showed that the proposed method can be used for robotic applications such as robotic kinematic calibration and 3D surfaces recognition and recording. For this last purpose the test rig is fitted on a robot arm that permits to the scanner device to ‘observe’ the 3D object from different and known positions.

1. INTRODUCTION

Laser scanning range sensors are widely used for high-precision, high-density three-dimensional (3D) reconstruction and inspection of the surface of physical objects [1]. The process typically involves planning a set of views, physically altering the relative object-sensor pose, taking scans, registering the acquired geometric data in a common coordinate frame of reference, and finally integrating range images into a non-redundant model [2]. Efficiencies could be achieved by automating or semi-automation this process.

The first aim of the research was to study an analysis procedure to elaborate image of laser in order to obtain 3-D object reconstruction, after an opportune system calibration. The technique has been automated by developing an interactive GUI to acquire and to elaborate data.

Reverse engineering is concerned with the problem of creating computer aided design (CAD) models of real objects by interpreting point data measured from their surfaces [2, 3]. For complex objects, it is important that the measuring device is free to move along arbitrary paths and make its measurements from suitable directions. This paper shows how a standard industrial robot with a laser profile scanner can be used to achieve that freedom. The system is planned to be part of a future automatic system for the Reverse Engineering of unknown objects. The system was designed around a test rig developed in our department by means of a commercial linear laser and a common web-cam; this system is moved by a revolute robotic arm.

1 Cesare Rossi: Tel.: (+39) 081-7683269; Fax: (+39) 081-2394165; E-mail: cesare.rossi@unina.it
2 Vincenzo Niola: Tel.: (+39) 081-7683269; Fax: (+39) 081-2394165; E-mail: vincenzo.niola@unina.it
2. EXPERIMENTAL PLATFORM

2.1 LASER SCANNER MODULE

Our rig, is based on a laser profile, that essentially consists in a line laser and a camera. The laser beam defines a “laser plane” and the part of the laser plane that lies in the image view of the camera is denoted the “scanning window”, figure 1. After an opportune system calibration it is possible to find a relation that permits to compute in the camera frame, the points coordinates of the scanning belt on the object surfaces, starting to its image coordinates. In this way it is possible to do a 3-D objects reconstruction by means of a laser beam.

The laser scanner device was realized by assembling a commercial linear laser and a common web-cam. The calibration operation provides to evaluate all parameters that are necessary to define the scanner module model.

2.2 ROBOT

In order to optimize the accuracy of the reconstruction resulting model, scanning should be adapted to the shape of the object. One way to do that is to use an industrial robot to move a laser profile scanner along curved paths.

The scanner laser module was mounted on a revolute robot with three d.o.f., designed and assembled in our Department, figure 2.
The robot serves as a measuring device to determine the scanning window position and orientation in 3D for each camera picture, with a great precision. All scan profiles captured during a scan sequence must be mapped to a common 3D coordinate system, and to do this, positional information from the robot were used [4]. Figure 3 shows the equipment at work.

The authors have developed a solution where the robot controller and scanner software work separately during a scan sequence that will be described in the following paragraphs.

3. EQUATIONS OF THE MODEL

3.1 MODEL OF THE LASER SCANNER

Camera observes the intersection between laser and object: laser line points in image frame, are the intersections between image plane and optical rays that pass trough intersection points between laser and object. In the figure 4 is shown a sketch of the laser-camera system.

For each point in the image plane, under the assumption of the pin hole camera model, it’s possible to write the follow relationship:

\[
\begin{align*}
  x_c &= (u - u_o) \delta v t \\
  y_c &= (v - v_o) \delta u t \\
  z_c &= ft
\end{align*}
\]  

(1)
With:

\((u_0,v_0)\): image frame coordinates of focal point projection in image plane;

\((\delta_u, \delta_v)\): fisical dimension of sensor pixel along direction \(u\) and \(v\);

\(f\): focal length.

For all the points in the image plane that represent the laser path namely the condition that they must belong to the laser plane as shown following:

\[
\begin{align*}
\begin{cases}
    x_v = \frac{(u - u_0) \delta_u z_v}{f}, \\
    y_v = \frac{(v - v_0) \delta_v z_v}{f}, \\
    Ax_v + By_v + Cz_v + D = 0
\end{cases}
\end{align*}
\]  

(2)

Where, considering a generic point \(\vec{p}_v\) in the laser plane and a vector \(\vec{v}\) orthogonal to the laser plane, whose components are \(A, B\) and \(C\), \(D\) is defined as follow:

\[
\vec{p}_v = \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 1 \\ A \\ B \end{bmatrix}, \quad D = -A\vec{x} - B\vec{y} - C\vec{z} 
\]  

(3)

From the system (2) it’s possible to write the coordinates of the points of the laser path in the camera frame:

\[
\begin{align*}
\begin{cases}
    x_v = \frac{(u - u_0) \delta_u z_v}{f}, \\
    y_v = \frac{(v - v_0) \delta_v z_v}{f}, \\
    z_v = \frac{-Df}{Af (u - u_0) \delta_u + Bf (v - v_0) \delta_v + Cf}
\end{cases}
\end{align*}
\]  

(4)

Equations (4) permits to compute in the camera frame, the points coordinates of the scanning belt on the object surfaces, starting to its image coordinates \((u,v)\), in pixel. In this way it is possible to do a 3-D objects reconstruction by means of a laser knife.

3.2 Model of the Laser Scanner on the Robot

When the laser scanner module is installed on the robot, figure 3, it is possible to use positional information from robot to determine the scanning window position and orientation in 3D.

Defining \([DH]\) as the transformation matrix between coordinates in the robot base frame \(0\) (the fixed one) and those in frame \(3\) (the one of the last link), figure 5, for the coordinates of a generic point \(P\) exists this relationship:

\[
\begin{bmatrix} P^0 \end{bmatrix} = [DH] \cdot \begin{bmatrix} P^3 \end{bmatrix}
\]  

(5)

The matrix \([DH]\) depends on 9 constant kinematic structure parameters, that are known, and 3 variable joints position parameters that are measurable by means of robot control system.
Knowing the transformation matrix $[^{T_i}T]$ between the camera frame and the frame of the robot last link, it’s possible to obtain a transformation matrix between the camera frame and the frame 0, figure 6.

$$\{P\}_c = \left[^{T_i}T\right] \cdot \left[^{DH}D\right]^{-1} \cdot \{P\}_0$$  \hfill (6)

By means of the equations (4) and (6), the relationship between image coordinates $(u,v)$ of the laser path and its coordinates in the robot base frame 0, is defined. By means of these equations, it is possible to reconstruct the 3D points in the robot base frame, of the intersection between the laser line and the object.

Robot positioning errors do not influence the 3D reconstruction, because each image is acquired and elaborated in a real robot position, that is known by means of robot encoders [5].

4. DATA CAPTURE AND REGISTRATION

The scanner video camera captures profiles from the surface of the object as 2D coordinates in the laser plane. During a scan sequence the laser scanner module is moved in order to capture object images from different sides and with different angles-shot according to the shape of the object.

In Matlab, an interactive GUI was developed in order to allows users to acquire and to elaborate data, figure 7. For each camera picture, along the scan path, the scanner derives a scan profile built up of point coordinates, in real-time.

The first step present in the GUI is the load of the calibration parameters that are composed by the laser scanner parameters and the matrix $[^{T_i}T]$. The second step is to filter the pixels of the laser path from the image, to do this, there are some regulations: the identification of the intensity of selected pixels, the calculus of the threshold and other regulations of the camera settings. The third step is to write the 3 joint position parameters of the robot in the window “position”, after this, clicking on the button “Image” the software save all the information, necessary for
the reconstruction, in the workspace of the Matlab. Clicking on the button “3D generation” the software calculates the 3D positions of the laser path in the robot base frame, and the result is shown on the GUI window.

When the scanning procedure is completed, the user can save images and relative robot position information in a file, save the cloud of points represent the surface of the test object and export the surface information in a file format that permits to load the data from the CAD software “CATIA”.

Besides it’s possible to load image information from a preview scanning procedure, this is useful for reconstruct the same laser path information using different calibration parameters.

![Figura 7:developed software](image)

5. EXPERIMENTAL RESULTS

The system has been tested before in a fixed robot position, to verify calibration and reconstruction procedures, then the shape of some components, was defined using robot to move laser scanner module. The test objects are shown in the figure 8.

![Figure 8: test objects](image)
In the figure 9 and 10 it’s possible to see a step of the procedure with the final results for the two test specimens.

Figure 9: elaboration procedure of the first test specimen

Figure 10: elaboration procedure of the second test specimen

With the use of the software CATIA it was possible to build the surface of the two test objects, in this way it was obtained the CAD model, this step of the 3D reconstruction method is a real reverse engineering application. The routine “Digitized shape editor” of the “CATIA” addresses digitalized data import, clean up, tessellation, cross sections, character line, shape and quality checking. In the figure 11 and 12 are shown the comparisons between the clouds of points and the respective surfaces for each object.

Figure 11: first test specimen results
6. CONCLUSION

The first results of a new method for real time shape acquisition with a laser scanner are presented. Although the test rig has been conceived just to validate the method (hence no high resolution cameras were adopted), the tests have showed encouraging results.

The test results also confirmed that it is possible to calibrate the intrinsic parameters of the video system, the position of the image plane and the laser plane in a given frame, all in the same time. Moreover the surface shapes were recognized and recorded with an appreciable accuracy.

The tests also showed that the proposed method can be used for robotic applications such as robotic kinematic calibration and 3D surfaces recognition and recording. For this last purpose the test rig was fitted on a robot arm that permitted to the scanner device to ‘observe’ the 3D object from different and known position.

Further tests will concern a detailed analysis of the sources of errors and the verification of the accuracy achieved. As far as the latter aspect is concerned, the authors believe that a better system for tracking the position of the robot arm could enhance accuracy.

Finally, the authors would like to point out that the solution proposed is relatively low cost, scalable and flexible. It is also suitable for applications other than RE, like robot control or inspection.

REFERENCES

Development of a Modular, Reconfigurable End Effector for the Plastics Industry: A Case Study

Michael A. Saliba*, Andrew Vella Zarb, and Jonathan C. Borg

Department of Industrial and Manufacturing Engineering
University of Malta
Msida, MSD 2080, Malta

ABSTRACT

A common problem in large manufacturing companies in the plastics industry is the need to handle mass produced injection-moulded items in the presence of a high degree of variety between batches. In this work, we address a problem being faced by such a company, where different batches of freshly-produced units need to be unloaded from the injection moulding machines. The products need to be grasped and relocated, and the runners disposed of. The number of identical units produced simultaneously by each single mould varies between two and sixteen, and the company makes use of around four hundred different moulds. We have developed a single modular end effector that can be easily reconfigured for the different moulds used by the company, and that also caters for future product designs. We present the detailed design of the new solution, including a function analysis derived from the existing process, a new product design specification, and a quality function deployment exercise. This is followed by the generation of a number of conceptual solutions to the problem, and the evaluation of the alternative concepts using morphological charts, failure mode and effect analysis, and a decision matrix. We present also a detailed embodiment design of the selected concept, and report on prototype development, evaluation and testing. Finally we discuss potential improvements through the use of the new system, and extrapolate the results of this work to more general applications where an end effector is required to carry out simultaneously several different but well defined functions in the presence of high variety.

1. INTRODUCTION

Manufacturing automation systems have traditionally been categorised into three distinct types, these being fixed, programmable, and flexible (see, for example, [1]). Fixed automation refers to a system that is custom designed for the production of a particular product, and is indicated where the production volume is high but where the product has little or no variety. Such a system has a number of benefits due to its dedicated nature, including relatively low cost and simplicity, however it has the major potential drawback of becoming obsolete if and when the product is discontinued. Programmable automation refers to a system in which the sequence of operations and motion parameters can be changed to accommodate different product batches, and is indicated in situations where production volumes are low but where there is relatively high product variety. Programmable automation systems are typically computerised numerically controlled (CNC) machines or industrial robots. While offering the significant advantage of being applicable to a variety of products, these systems normally require a high investment cost, and normally require a significant amount of set-up and programming time between batches of different products. Flexible automation refers to a system that can handle a variety of products without set-up or reprogramming. Such systems are widely regarded as representing an ultimate achievement in automation, however they are not without drawbacks. Generally, they are relatively complex systems that come at a substantial cost, and moreover the amount of product variety that they can process is often quite limited.

Today’s market is characterised by tough competition that continually forces manufacturers to push down their costs, coupled with a discerning customer who constantly demands a higher level of variety. This scenario has led to the search for new automation systems that combine advantages of the three traditional approaches, while mitigating the disadvantages of each. The ideal automation system would be cheap and simple to implement, while at the same time would be able to handle a high degree of variety with minimal set-up time between batches of different...
products. This search has led to the development of a new paradigm in manufacturing automation, that of reconfigurable automation [2], [3], [4], [5], [6]. The basic concept of this approach is to develop an automation system that is very modular in nature, and where the constituent modules can be quickly interchanged, exchanged, or otherwise reconfigured to handle different products. Once the two fundamental problems of module definition and of module interfacing have been solved, these systems can become quite inexpensive and simple to implement, have the capability to handle a significant amount of variety, and can be reconfigured (normally manually) relatively fast.

A very common requirement in manufacturing automation systems is that of grasping and handling objects, and as such, significant effort is very often made to maximise the effectiveness of this function within the automation system (e.g., [7], [8]). Where it is required to find a relatively low cost solution to dealing with the gripping of objects in the presence of variety, the use of a reconfigurable gripping device may be indicated. This is often done by developing a gripper in which the positions of some or all of the fingers can be adjusted to accommodate each particular batch of objects before the start of each production run (e.g. [9], [10]).

In this work, we address a handling problem related to the plastics injection moulding industry, namely the transfer of freshly produced parts from a mould to a conveyor, on which the parts later move for further processing. The issue is addressed in the context of an actual case study, in which a company is seeking to improve an already automated process. As shall be seen below, the problem at hand has a number of distinctive features. Apart from a high degree of variation in part size and shape (approximately four hundred different products) that need to be handled, the number and positions of idential parts manufactured simultaneously within each mould also differs. Furthermore, in addition to grasping the parts, the parts handling device has to cut and dispose of the runners and sprues associated with the injection moulding process. The grasped parts need to be released on to the conveyor in a highly ordered manner, in order to be ready for further automated processing. The main general objectives that need to be achieved are a reduction in the cost associated with the development and set-up of the gripping device whenever a new product is introduced by the company; and a reduction in the required storage space for the gripping devices and associated sub-components that are not in use. For the reasons indicated above, the approach selected to solve this problem is the design and development of a modular, reconfigurable end effector.

2. PROBLEM DEFINITION

2.1. THE COMPANY

The company involved in this work is privately-owned and deals with the manufacture of high quality plastic packaging. The plastic products go from raw material to the finished packaging product in-house. The cases and containers produced are then sold to client companies that complete the production process and distribute the products to retailers, for eventual sale to the end users. All the cases produced by the company have to meet very high standards of quality on both an aesthetic and a functional basis. To produce such high quality products, the production equipment used is centred on quality, and the margin of acceptable defects is very small.

To retain its competitive edge the company is always seeking to improve its manufacturing system. These improvements can range from the installation of new machinery to simply altering a sequence of production operations using the current installation. These changes can often work out to be costly and then become counterproductive as the product price may increase, thus reducing the economic competitiveness. However product quality can be improved without increasing cost, by investing in automation and lean manufacturing techniques. These techniques strive to reduce waste and thus increase profitability by reducing the overall product cost.

In light of this, the company is constantly seeking to reduce waste (both material and time) in various stages of the production operations. This is what has led to the formulation of this project.

2.2. THE CURRENT PROCESS

The operation under consideration involves the unloading of the injection moulded parts from the open mould, the disposal of the runners, and the transfer of the parts to a conveyor system. This operation is already automated using Cartesian robots, but the company has a different end effector associated with each mould, to grasp and relocate the units and to cut and dispose of the runners. The number of identical units produced simultaneously by each single mould varies between two and sixteen. The company makes use of around four hundred customized end effectors, with an associated penalty in storage, handling, and set-up costs. The company designs and develops a
new end effector whenever a new product is commissioned. The maximum combined weight of the parts to be transferred is of about 3N.

An example of an end effector that is currently used for this operation is shown in Figure 1(a). The end effectors, or robot plates, make use of vacuum grippers (Figure 2(b)) to grasp the moulded parts, and jaw grippers (Figure 1(c)) to grasp and snap off the runners. For the example given, the mould produces four units per cycle, with the positions of the finished units within the open mould corresponding to the positions of the four vacuum grippers on the robot plate, which would have been custom built for this particular mould.

![Figure 1: (a) Robot plate, (b) Vacuum grippers, and (c) Jaw gripper](a) (b) (c)

The work is batch oriented. This means that a particular product is produced for a limited period of time to fulfill a customer’s needs. Once the pre-ordered number of products have been produced, the product is discontinued for an amount of time which may vary. Every time a batch has been completed, the tooling used is removed from the injection machine, the particular robot plate is removed from the robot wrist mounting plate, and the controller programs are saved and removed from the controller. Once the injection machine is needed for a new batch (of the same product or even different products) the tool is once again mounted on the injection machine platen if already manufactured, its respective robot plate is mounted or custom made, and the relevant controller program is either re-loaded or written from scratch.

Every mould tool requires a particular robot plate. The number of mould tools used already exceeds four hundred, and new products are continually developed. The customized robot plates are not only expensive to produce, but are also expensive to store due to the large number of plates kept.

### 2.3 Function Analysis

The parts are gripped from the open mould, in the vertical position. In this position the retaining force of the vacuum grippers is at its lowest, being exposed to a shearing force by the weight of the part. The robot plate is then raised above the open mould, still in the vertical position. Once clear of the mould, the wrist mounting plate then rotates downwards, thus positioning the gripped parts directly below the robot plate. Here the vacuum gripper retaining force is higher since the weight of the parts is acting directly against the vacuum force applied. A vertical and horizontal translation by the robot then positions the robot plate over a disposal bin. Here the jaw grippers are opened, releasing the runners which fall into a disposal bin. The robot plate is then translated horizontally and over the conveyor belt. Once in position, the robot plate is then lowered, gently placing the products onto the conveyor. The cycle time for this operation, including all mould and robot motion, is of nineteen seconds. The problem boundary, based on the transparent box approach, is shown in Figure 2.

### 2.4 Product Design Specification and Quality Function Deployment

A product design specification (PDS) for the end effector was drawn up, outlining the customer’s requirements in terms of demands and wishes. With regard to the working environment, a demand was that the end effectors or robot plates needed to be able to work in proximity to hot mould tools. To satisfy performance requirements, the device needed to perform the basic part transfer operation; maintain the existing level of positioning accuracy; not damage the parts; be adequate for future products from the same part family; be able to handle hot, freshly moulded parts; and be easy to set up or reconfigure. The device needed to have a service life that would allow it to process one entire batch (approximately 50,000 units) as an absolute minimum. The device also had to be compatible with the existing bulk machinery. With regard to kinematics, the device would have to withstand high accelerations. The device weight was limited to a maximum of 5 kg, and size to 250 mm, due to robot payload and mould gap constraints.
During the disposal phase, the device needed to not harm the environment. To satisfy materials requirements, the device had to not affect the facilities and products; it had to be lightweight; it had to resist warping and deformation; and it had to be safe for workers to handle.

The design wishes included an ability to handle as large a number of batches as possible within its lifetime, and to have consistent performance characteristics throughout its life; to be inexpensive to develop and to have low running costs; to make use of readily available forms of energy; to be small and easy to store when not in use; to be easy and inexpensive to maintain; to be able to be fully or partially re-usable at the end of life; and to be made of readily available materials.

A quality function deployment (QFD) chart was drawn up to relate the company’s requirements and preferences to specific technical parameters. This is shown in Figure 3.

3. CONCEPT GENERATION

A morphological chart of candidate subsystems that would perform each of the required functions of the device was drawn up. A number of alternative concepts were generated for each element in the morphological chart. The concepts that were generated for the working principle are shown in Figure 4.

In the concept shown in Figure 4(a), the matrix of pins conforms to the object to be grasped, and the extended pins are then compressed against the object by a surrounding inflatable bladder. The concept of Figure 4(b) consists of a matrix of suction grippers, which would be used in conjunction with a customized blanking plate. In both these concepts the runners are disposed of in a separate process. In Figure 4(c), the conceptual solution consists of a flexible, soft polymer pad with perforations that can be connected to a vacuum pump. Runner removal can be performed using a high powered air jet. The concept of Figure 4(d) consists of a reconfigurable plug board to which suction and jaw grippers can be attached as required for each mould. In Figure 4(e), the conceptual solution consists of individual suction and jaw grippers that can be inserted in various positions along horizontal rails. The rails themselves can be attached to cross-rails for further position setting.

4. CONCEPT EVALUATION AND SELECTION

Concept evaluation was carried out using two design tools, namely failure mode and effect analysis (FMEA) in order to reveal potential failures of each conceptual solution, and a decision matrix to compare different solutions.

The first step in constructing a decision matrix was the establishment of the evaluation criteria. These were mainly based on the requirements specified in the Product Design Specification list and also on the constraints around the developed product. The criteria included several technical, economic and also safety factors, thus giving a broad spectrum of criteria on which the concepts could be evaluated. The factors chosen were those which are decision-relevant, as well as general constraints. The factors are also relatively independent of one another.
Not all the evaluation criteria have the same importance to the overall goal, since some criteria contribute more to the overall satisfaction of the customer requirements than others. The relative importance of each of the criteria is shown by the value of a weighting factor associated with each particular criterion. A weighting scale from 1 to 10 was used, 10 being the most important, and 1 being the least. The weighting values were determined from detailed discussions with the customer. Criteria which were given a weighting value of 10 were considered to be indispensable, and thus if a concept failed to satisfy these criteria, the concept in question was not considered an adequate solution to the overall project problem. The computed decision matrix is given in Table 1. The highest score in this exercise was obtained by Concept 5, the reconfigurable rails concept.

The FMEA exercise for the reconfigurable rails concept showed that the most likely potential failure modes of this concept could be countered with simple maintenance routines and good overall design. This concept was therefore selected for further development.
(a) Pin-box  
(b) Suction gripper matrix and blanking plate

(c) Perforated polymer pad  
(d) Plug board  
(e) Reconfigurable rails

Figure 4: Generated concepts for the main working principle

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No part damage</td>
<td>10</td>
<td>20</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>High shape flexibility</td>
<td>10</td>
<td>80</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>High mould layout flexibility</td>
<td>10</td>
<td>70</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>Low set-up time</td>
<td>9</td>
<td>81</td>
<td>72</td>
<td>90</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td>Low running cost</td>
<td>7</td>
<td>21</td>
<td>49</td>
<td>56</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td>Low initial cost</td>
<td>3</td>
<td>12</td>
<td>21</td>
<td>9</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td>High accuracy</td>
<td>8</td>
<td>40</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td>Ease of development</td>
<td>9</td>
<td>36</td>
<td>81</td>
<td>18</td>
<td>54</td>
<td>90</td>
</tr>
<tr>
<td>Low storage usage</td>
<td>4</td>
<td>40</td>
<td>28</td>
<td>40</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Low modification of robot/tool</td>
<td>7</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Long lifespan</td>
<td>6</td>
<td>48</td>
<td>54</td>
<td>42</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>High maintainability</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>20</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td>Easy to re-use/recycle components</td>
<td>2</td>
<td>4</td>
<td>16</td>
<td>6</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total Weighted Score</strong></td>
<td></td>
<td><strong>509</strong></td>
<td><strong>688</strong></td>
<td><strong>644</strong></td>
<td><strong>722</strong></td>
<td><strong>817</strong></td>
</tr>
<tr>
<td><strong>Percentage Score</strong></td>
<td></td>
<td><strong>15</strong></td>
<td><strong>20.4</strong></td>
<td><strong>19.1</strong></td>
<td><strong>21.3</strong></td>
<td><strong>24.2</strong></td>
</tr>
</tbody>
</table>

5. EMBODIMENT DESIGN, FABRICATION AND TESTING

After the solution concept had been selected, a new morphological chart was drawn up to evaluate and select the best methods to achieve the required functions of the prototype. Two key parameters that were analyzed were the locking mechanism between the grippers and the mounting rails, and the sliding mechanism between the mounting rails and the cross rails.

For the locking rail mechanism, the generated concepts included magnetic locks, screw friction locks, bolt locks, and spring loaded pin in hole locks. These candidate solutions were evaluated in terms of flexibility, set-up time, positional stability, cost, and availability. The selected concept for this function was the spring loaded pin in hole lock.
lock, due to its high position stability, and low cost and set-up time. A CAD drawing of the gripper fixture design is shown in Figure 5(a). A similar mechanism was selected for the movement and locking of the mounting rails relative to the cross rails, and a CAD drawing of the fully assembled system is shown in Figure 5(b).

![Figure 5: (a) gripper fixture design, (b) reconfigurable rail design](image)

After the detailed designs of the reconfigurable gripper were completed, a prototype was fabricated in order to test the reconfiguration and locking abilities of the system. For the prototype, the rails were fabricated out of mild steel box section, and the gripper fixtures were fabricated out of aluminium. Details of the prototype can be seen in Figure 6.

![Figure 6: (a) gripper fixture, (b) and (c) reconfigurable rail](image)

All of the sliding and locking functions of the prototype have been tested, and have been found to function well under the expected loads. During testing to failure, failure of the locking mechanisms occurred at between 6 and 8 kg loading. This load is well above the maximum weight of the parts to be handled, and exceeds also the payload capacity of the robots (see sections 2.2 and 2.4). Testing on the production equipment is scheduled to commence shortly.

6. DISCUSSION

This work has addressed two important issues in industrial automation, as discussed briefly in this section. The first issue relates to the critical decision that often needs to be made between flexibility and reconfigurability in manufacturing automation. In the current application, if a flexible solution had been selected, the gripper may have been able to handle the different sets of parts from all 400 moulds without human intervention between batches. The key question here however, is whether the greatly increased complexity and cost of such a gripping device would be justified in an application where large batches of approximately 50,000 identical units are being processed. The answer to this question, in our opinion, is negative. By selecting instead a reconfigurable solution, it is seen that, due to the large batch size, the added penalty incurred in reconfiguration time is in fact negligible when compared to the production time for each batch, and is well offset by gains in simplicity, reliability and lower cost.
The second issue relates to the fairly common problem, particularly in the plastics industry, of needing to grasp different sets of different objects in batch-wise fashion from the same working area. In this application, each set of objects corresponds to a particular mould, and in each set there are two types of objects to be grasped – the products and the runners – requiring two very different types of grippers. We have addressed this problem using two of the basic concepts of reconfigurability – modularity and adjustability. Through modularity, we are able to attach different types of grippers (in this case, suction-type and jaw-type) as required, and through adjustability we are able to place these grippers at the required locations. It is further noted that a key issue in modularity is the design of an appropriate interface between modules, and in our work this was achieved through the development of an appropriate setting and locking mechanism for the gripper fixtures to the rails. The required adjustability was achieved through the use of the sliding rails concept.

In the case study addressed in this work, the company is expected to achieve significant savings in costs related to the introduction of new products, since it will no longer need to develop new robot plates. The company is also expected to achieve significant reduction in the storage space need to store the handling equipment when not in use.

7. Conclusion

In this work, the design and prototyping of a reconfigurable end effector for the plastics industry has been presented. The developed concept has applications where different batches of products need to be unloaded from plastic injection-moulding machines during production, and where the number and location of products may differ between moulds. The end effector or so-called runners and sprues associated with the moulding process. A formal design approach has been adopted in the development of the solution, and this has led to a solution, from several alternatives, that best satisfies the various requirements of the customer. The developed solution may have further applications in automation outside of the plastics industry, in situations where a number of objects need to be grasped simultaneously using different types of grippers, in an environment of high variety between batches.

References

Perceptions and Problems Related to the Use of Manufacturing Automation in a Small, Geographically Isolated Economy

Dawn Zammit\(^1\), Sandro Azzopardi\(^1\), Michael A. Saliba\(^1\)*, and Martin M. Zammit\(^2\)

\(^1\) Department of Industrial and Manufacturing Engineering
\(^2\) Department of Management

University of Malta
Msida, Malta

ABSTRACT

In small, geographically isolated economies, it is possible to find a lack of appropriate support for the implementation of state-of-the-art technologies and methodologies for manufacturing automation, thus compromising the efficiency and competitiveness of the manufacturing sector in the global marketplace. In this work, we carry out a critical investigation of the use of automation within the manufacturing sector of such an economy, using the island nation of Malta as a case study. We focus mainly on high value-added manufacturing, characterized by the need for high product variety and relatively low product quantities, and triggering the need for versatility in production facilities, potentially in the form of reconfigurable production systems. We have developed a set of detailed questions, in the form of a survey questionnaire, which we pose to a sample of companies from various sub-sectors, and of various sizes, within the Maltese manufacturing sector. Respondents answer the questions in the setting of a semi-structured interview, following an extensive familiarization tour of the company facilities by the interviewers. We present the results, as well as their analysis and interpretation, from our first sample of thirty manufacturing firms, ranging in size from micro to large, and taken mainly from the electronics, medical, pharmaceuticals, plastics, food and beverage, chemicals, and furniture industries.

1. INTRODUCTION

In the past decade, various authors have discussed the emerging challenges related to the continuous struggle faced by manufacturing companies to improve efficiency and effectiveness [1]-[5]. These challenges affect different manufacturing sectors to varying extents, however in general they continue to become more stringent, for a number of reasons. Firstly, customers are expecting a much larger amount of product variety and increased quality. Secondly, as time passes, there is continuous progress in technology, and as a result, products are constantly evolving so as to incorporate new technologies. Manufacturing systems and processes are also being affected with this advancement of technology due to the development of new techniques, processes and equipment. The introduction of regulations is already having an effect on the manufacturing industry, both due to regulations affecting the product, as well due to regulations affecting the production of products or the processes. Another existing challenge concerns the shift to an ever-expanding globalised market, with manufacturing companies no longer having to compete only against local companies, but nowadays having to compete against distant and/or foreign companies as well.

However, such studies have generally been based on relatively large countries and economies, or else been based on very general overall situations. The specific aim of our work is to carry out a critical investigation of the manufacturing situation within a small and geographically isolated economy, with a focus on the implementation of industrial automation, and to identify and address problems that exist. So solutions to identified problems will be developed with a focus on the reconfigurable manufacturing systems paradigm. This paradigm was developed in the early 1990s and early 21st century, and various authors have discussed Reconfigurable Manufacturing Systems (RMSs) [2], [3], [5], Reconfigurable Equipment [6], and Reconfigurable Machine Tools [7]-[9]. Such systems bridge the gap between Dedicated Manufacturing Systems and Flexible Manufacturing Systems, and combine the advantages of both approaches. RMSs are built around a part family, or part families, allowing product variants within the respective part family/families to be manufactured through a

* Corresponding author: Tel.: (00356) 2340-2924; Fax: (00356) 2134-3577; E-mail: michael.saliba@um.edu.mt
rapid and relatively simple reconfiguration of the equipment. Reconfigurable equipment has the potential of substituting, exchanging, adding, removing, and/or modifying specific modules to change a new existing RMS configuration into a new configuration with different capabilities.

This paper focuses on the first part of our overall project. The primary targets of this specific study are to investigate the perceptions and problems relating to manufacturing, with a focus on automation, that are experienced specifically by companies operating within a small and geographically isolated economy. As a case study, the current situation in the Maltese Islands has been taken.

Malta is an island nation, and European Union (EU) member state, in the middle of the Mediterranean Sea. It has a population of 400,000 and a gross domestic product (GDP) of about €5.7 billion ($7.4 billion) [10]. The manufacturing sector accounts for about one fifth of the GDP. The Maltese economy has experienced a transition in the last decade whereby the traditional low technology mainstream manufacturing sector experienced difficulties due to its heavy reliance on manual labour. Therefore, Malta’s focus has been directed towards high value-added manufacturing, in an attempt to enable industry to achieve productivity growth to engender prosperity and wealth. Analysis has shown that the European manufacturing industry consists in large majority of SMEs (Small and Medium Enterprises), with over 99% of companies and 58% of manufacturing employment falling within this group. In fact, these SMEs generate the majority of new and innovative products, providing some three-quarters of EU exports. Malta’s economy as in the case of other EU countries is characterized by a predominance of SMEs. These businesses are not only fundamental to the on-going economic activity, but are also a complementing arm of industry in general since these would supply the larger companies making up the industry.

This study therefore has two objectives with respect to the Maltese manufacturing sector: an information gathering objective, and a problem identification objective, with the particular focus being automation in high value-added product manufacturing. The results of this study, although specific to the Maltese manufacturing environment, may also be widely relevant to other pocket economies within the EU, as well as within the rest of the industrialized world.

2. PRELIMINARY GUIDELINES

To thoroughly achieve these two primary objectives of this study it is required to carry out an extensive investigation of the local scenario through the gathering of information about various issues relating to manufacturing and industrial automation. The companies that have been investigated so far have been chosen at random from an exhaustive list of manufacturing companies compiled from databases taken from various sources.

<table>
<thead>
<tr>
<th>Sectors with Priority 1</th>
<th>Sectors with Priority 2</th>
<th>Sectors with Priority 3</th>
<th>Sectors with Priority 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics</td>
<td>Beverages</td>
<td>Meat Processing</td>
<td>Construction</td>
</tr>
<tr>
<td>Medical</td>
<td>Chemicals</td>
<td>Paper</td>
<td>Explosives</td>
</tr>
<tr>
<td>Pharmaceuticals</td>
<td>Food</td>
<td>Printing</td>
<td>Metal-work</td>
</tr>
<tr>
<td>Plastic-ware</td>
<td>Furniture</td>
<td>Tobacco</td>
<td>Mineral</td>
</tr>
<tr>
<td>Glass</td>
<td>Wax</td>
<td>Recycling</td>
<td>Shipping</td>
</tr>
<tr>
<td>Textiles</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The companies within this list were categorized according to the nature of the products produced within each company, e.g., medical, pharmaceutical, furniture, chemical, etc. Twenty-one manufacturing sub-sectors were outlined. These sub-sectors were put in a priority list as outlined in Table 1 in terms of the nature of the products manufactured. Perceived high-volume discrete product manufacture and consumer product manufacture were given a high priority (in this case level 1 and level 2), whilst products that tend to be one-offs, such as infrastructure, and high-volume production of the same product, such as paper manufacture, were given lower priorities (in this case level 3 and level 4). This study aims to investigate the high priority manufacturing sectors.

Apart from the manufacturing sector within which a particular company operates, another important factor that may lead to characteristic problems is the actual size of the company in terms of the total number of employees. Each category has thus been further subdivided into four sub-categories according to the number of employees within each manufacturing company, namely: micro-sized companies, having less than 10 employees and with a turnover less than €2 million; small-sized companies, having less than 50 employees and with a turnover less than
€10 million; medium-sized companies, having less than 250 employees and with a turnover less than €50 million; and large companies, having more than 250 employees, as per the definitions in [11].

3. SURVEY DEVELOPMENT

3.1 SURVEY FORMULATION

As mentioned in the previous section, this study has two main objectives; a critical investigation, and a problem identification objective. The gathering of information is mainly through a detailed survey, consisting of two main sections. The first section contains general questions relating to the company and the manufacturing approaches and means implemented. This first section is intended to be answered by all respondents. The second section focuses on automated manufacturing, and is therefore intended for respondents that already implement automated machinery.

The first section of the survey starts off with general questions related to the company being investigated, such as a description of the manufactured products, company size, as well as the number of employees. As mentioned above, certain characteristics may arise that can be correlated to each of these separate factors. Therefore, such questions were required to position the company in the right categories, mainly with respect to sector and size, so as to be able to analyse trends of these separate categories.

As indicated above product lifetimes are constantly decreasing, that is the time between the initial stages of production to the time when the product is not manufactured any more. This being a very important driver for the subsequent manufacturing approach implemented within a company, there was a need for a question related to the lifetime of the products being produced by the company being interviewed.

The next questions tackle specifically the manufacturing approach and processes currently being made use of within the production facility of the company being interviewed. In this specific study, the broad categorisation of manufacturing approaches has been taken to be manual, semi-automated and fully automated. Manual processes are processes made up of mostly manual labour. A semi-automated manufacturing approach refers to operations being split between manual and automated, whilst fully automated processes refer to processes run fully autonomously or with minimal human intervention. Respondents could choose any of these approaches; or multiple approaches if different approaches are undertaken for different manufacturing processes. The reasons given by the respondents for their choice of manufacturing approaches were also requested. Some potential reasons that were identified beforehand, and given to the respondents to choose from, include product variety, production volumes, and product/process complexity. In addition, respondents were also questioned about specific production processes implemented on their shop-floors, and whether such processes are fully manual, semi-automated or fully automated.

In order to determine the use, or the lack of use, of automation equipment, it was considered necessary to inquire about any perceived problems towards the implementation of automation equipment and industrial stand-alone robot arms. To gather with this information, other questions that tackle issues relating to the willingness to invest in automation in relation to customer pressure and/or to competitor facilities were also included in the survey.

The second section of the survey is more focused on automation and its implementation, and as mentioned, is reserved only for respondents that currently already employ some level of automation. The first question in this section relates to the identification of equipment that generally is implemented for automated manufacturing. Respondents are given a list of such automated equipment from which they have to tick any already implemented equipment. The list includes Computer Aided Design (CAD), dedicated automation equipment, specialised automation equipment, Computer Numerically Controlled (CNC) machining equipment, pneumatic/hydraulic devices, flexible or reconfigurable automation, and machine vision. This is required so as to determine what facilities are currently being used the most in automation. In addition, it was deemed useful to understand where the design and manufacture of such automation equipment is undertaken. Common approaches include companies designing and manufacturing their equipment in-house, with others subcontracting it to local as well as foreign mainland equipment manufacturers. In the case of certain subsidiary companies, the technology, processes, and even any required equipment may be directly adopted from the respective mother/sister companies. As a result, all these options were included from beforehand in the survey. For each option chosen, the respondents were also required to give the reasons for their choice. Some possible reasons that were included from beforehand in the survey include quality and available expertise; however respondents were encouraged to add to the list.
The next set of questions, relate to how the companies react with respect to the implementation of automation equipment when a completely new product is to be introduced to the existent product portfolio. This would be useful to determine whether a new system would be opted for straightaway or whether the existent system would be considered for the manufacturing of the new product thus resulting in reconfigurability of the automation equipment.

Besides reconfigurability of automation equipment, this study is also interested in flexible automation and whether this specific philosophy of manufacturing is actually implemented or not. A number of questions were introduced in order to determine whether equipment with extra in-built capabilities would be considered even though at that point in time, there would be no use for all its capabilities. Also, it was deemed useful to determine whether if this is opted for, the equipment would actually later be used to its full potential or not.

The next questions address the utilization of specific equipment for various different products, with a focus on reconfigurability and modularity. First, the issue of whether the current production equipment is to a particular extent modular or not was tackled. Following this, respondents were asked whether the equipment utilized for the manufacturing of a particular product is re-used or reconfigured when the lifetime of the respective product reaches an end. Subsequently, the frequency of “engineering changes” on an existing production system was requested, with the term “engineering changes” being used in his particular study referring only to changes that require a certain degree of new engineering input, and hence not including work order changes. This was followed by the reasons causing these changes, and a list was a gain made available to the respondents. The list included regulations and/or customer requests, quality requirements, capacity requirements among others. Finally, the typical extent of changes and the time taken to carry out the changes were asked. These questions would be required to determine the degree of changes and thus the requirement to adapt quickly to new set-ups.

The final question of the survey focuses on the relationship between product design and product manufacture, focusing on the extent to which the product design may be affected by the manufacturing set-up itself. Product design is one of the initial stages of the product life time, with manufacturing following at a later stage. This question would give an indication on how the product design may affect the modularity or reconfigurability of the production facilities used and the extent to which the product design may be affected in order to automate the process on which the product will be manufactured.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Micro</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beverages</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chemical</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Electronics</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Food</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Glass</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Medical</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Pharmaceuticals</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Plastic-ware</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Textiles</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Woodworks</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>30</td>
</tr>
</tbody>
</table>

### 3.2 Survey Implementation

Following a number of test runs, it was decided that the methodology adopted for the gathering of information through the survey would consist of conducting a semi-structured interview with an appropriate company representative(s). This company representative(s) ideally had to be very familiar with the production activities within the company, such as the head of manufacturing department or operations manager. The semi-structured interview included a physical tour within the manufacturing facility of the company and subsequently a discussion on the survey questions. The discussion as well as the tour proved to be a valuable source of information towards the completion of the survey questions. In fact, this was confirmed during the test runs that were conducted. Even though lengthy, this methodology improved consistency in the answers given by different respondents due to the fact that, the replies to the questions asked were always recorded by the same person. To date, thirty interviews have been conducted from the sub-sectors with priority weighting 1 and 2. The breakdown of the companies interviewed by sector and size can be seen in table 2.
4. PRELIMINARY RESULTS

This section discusses the results obtained from the preliminary analysis of the first sample of thirty surveys performed. However, it is pointed out that this analysis is still a general one due to the relatively small number of surveys performed within each sub-sector. The trends within each particular sub-sector and size category require more surveys to be performed.

4.1. PRODUCT LIFETIME

It was interesting to note that approximately half of the companies answered that their product has a lifetime of more than 5 years as illustrated in Figure 1. Another twenty percent produce products with a lifetime between two and five years. The rest of the respondents have products with lifetime of up to 24 months. Thus a trend can already be seen, with the majority of cases being companies having quite lengthy product lifetimes, which contrasts with some descriptions of current trends found in literature. Through various discussions with manufacturing personnel from such companies, a common reason for this behaviour was that the companies are subsidiaries having foreign mother companies, with the trend being that the mother companies move those products that have already reached their maturity stage to the plant in the small country. Newer products are possibly manufactured nearer to the customers, since they would still be under development. Since products that have already reached their maturity stage, in general, tend to start decreasing in volume, the companies in the small countries would end up producing an increasing variety of batches/types of products, with each of these having a relatively small volume. This would result in having a good number of products with relatively low volumes to be produced for a considerable length of time.

![Figure 1 - Product lifetimes against percentage number of respondents](image)

4.2. MANUFACTURING PROCESSES

Having a large variety but low volume products portfolio was given as one of the main reasons why fully manual manufacturing processes are still adopted by many of the respondents, totaling forty percent of the survey respondents. These respondents commented that the utilization of manual processes is due to requiring various changeovers in the production setup, and therefore necessitating a high degree of flexibility in their processes so as to be able to quickly adapt the process to all the product variations required. However, another primary reason for the adoption of manual labour is the complexity of mechanical assembly, with the most common manual processes being material transfer, packaging, and inspection.

Semi-automated processes were implemented when the requirements of quality and efficiency were deemed to be the primary manufacturing issues. In fact, this was the case in forty percent of the interviewed companies. In such cases, any mechanized processes such as injection moulding are also included. Other techniques include printing and cleaning processes. The semi-automated processes are also characterized by higher volumes. However, it is important to note that the responses noted that their clients do not generally impose the requirement of automating the processes.

Fully automated processes are only employed by about one fifth of the interviewed companies. The reason given for implementing fully automated processes is mainly due to high production volumes. The most common fully
automated manufacturing processes include soldering processes, especially in the electronics industry, as well as printing processes.

4.3. IMPLEMENTATION OF AUTOMATED EQUIPMENT

With the implementation of automation being relatively limited, it is important to investigate whether there are some perceived problems for the implementation of automation. A common and significant problem that has been encountered in practically all companies is the lack of time to study opportunities. A reason for this may be that in small and geographically isolated countries, most often all companies are relatively small. Even companies that are regarded locally to be large, in fact are still relatively small when compared to mainland companies. This results in these smaller companies having to compete with larger competitors while seeing everyday routine non-manufacturing tasks as well, such as maintenance works. These everyday non-manufacturing tasks would make it difficult for them to study opportunities to enhance their productivity. In addition, the exposure to newly developed technologies through relevant industrial shows, exhibitions, and fairs is minimal, due to the relative isolation of the economy.

Besides the above discussed reasons, there are other perceptions that hinder the use of automation equipment. Continuing from the lack of exposure to manufacturing technologies, these include the perception that such equipment is not appropriate for the products or for the volumes being produced. This is especially identified when automation equipment such as stand-alone industrial robots were discussed. Another high rating perception is that the equipment and its relevant required implementation are too expensive. Such respondents argue that it all depends on the volumes being produced. With the already identified low production volumes, such investments are regarded as not being justifiable.

4.4. DESIGN AND MANUFACTURE OF MANUFACTURING AUTOMATION EQUIPMENT

When automation equipment is actually implemented, the most common approaches include making use of dedicated and specialized automation equipment, flexible and reconfigurable equipment, as well as machine vision. The common approach is for the design of such equipment to be outsourced to specialized manufacturers outside of the country; however modifications and adaptations are usually done in-house. This is due to the fact that for such complex manufacturing equipment the knowledge of complex processes and technologies is considered to be in the hands of developers who are specialized within that particular manufacturing field, and thus are readily equipped with the knowledge and technologies required. Therefore, local manufacturers outsource this equipment with the perception of obtaining high quality equipment. The manufacturing of such equipment is also mostly left in the hands of these foreign based industrial equipment manufacturers. However, when it comes to the development of relatively simple equipment or a slight modification to equipment, the common approach is to outsource such tasks to local manufacturers or else they are performed in-house, with both these approaches opted for due to financial reasons.

4.5. NATURE OF THE PRODUCTS AND INDUSTRIAL MANUFACTURING EQUIPMENT

It was observed that many of the investigated companies specialize in the production of part families, with the products that they manufacture being to a certain extent similar to each other in nature. In fact, in most cases, even new products that would be added to the current product portfolio would also be similar in nature to those already existing or to past products. Therefore, it would be possible to include them in already existing setups without major changes. In fact, when discussing this issue, most of the survey respondents said that when new products are to be introduced, these are in fact added on to the already existing production lines. In fact, when investing in industrial manufacturing equipment, the common approach is to invest in equipment having extra in-built capabilities. The approach taken is to utilize this equipment to manufacture the product portfolio, which is as broad as possible mentioned, within each manufacturing company. These considerations can be regarded to be a broad product family. Still, respondents pointed out that many times, the extra in-built capabilities invested in when buying the equipment are still not all made use of. When it comes to the end-of-life of a current product, these trends are reflected on the approaches taken with regards the manufacturing equipment that was already utilized for that particular product. This is illustrated in Figure 2, which illustrates the most common actions taken for the manufacturing equipment for new products.

4.6. CHANGES TO EQUIPMENT

It has already been established that the lifetime of the products manufactured by the companies interviewed, is, in general, relatively long. This is together with the fact that the products would be past their development stage, would mean that no alterations of “engineering changes” would be required in the processing facilities. In fact as opposed to work order changes which occur on a daily basis due to the various different products manufactured, the
survey showed that such engineering changes occur very rarely, with these changes occurring every few years. The main reasons for these changes when they occur include requirements for quality improvements, new implemented regulations, new products adaptation, as well as capacity changes that are beyond what was catered for initially. The typical extents of these changes are minimal hardware and/or software changes; with the time to carry out such changes ranging from only a few hours to a few weeks.

4.7. PRODUCT DESIGN

Another interesting point that arose from this study is that no product design activities are present in most of the investigated companies, with the main reason being that the majority of respondents are subsidiaries to larger companies. In these cases the trend is that the mother company develops the product design, with this design being handed over to the local companies. These local subsidiaries would therefore have to manufacture the already designed products. As was discussed in a previous section, these products tend to also have already reached their maturity stage, and therefore would be well past their developmental process. This further limits any possible changes to the product design.

5. DISCUSSION

Many of the key characteristics of manufacturing companies that were identified in this study, with regards to manufacturing and industrial automation equipment, may be relevant to other small and/or geographically isolated economies. One of the key points is that most of the thirty companies investigated have a product portfolio composed of broad product families, each of these having a large amount of products which despite having differences amongst themselves, can still be categorized and grouped together due to a high percentage of similarities. During the investigation, this was mostly encountered in micro-sized companies. The end result would be that such companies opt for fully manual manufacturing processes with the perceptions that there is no other feasible way. The lack of industrial fairs and locally based industrial equipment manufacturers continue to reinforce this perception. It is clear that manufacturing companies operating within such a pocket economy can not stay competitive for long without the exposure to new and improved manufacturing approaches, ways, and means. The survey has revealed that companies that do attend industrial fairs tend to be much more open to innovative and advanced manufacturing methods. This implies that such exposure may be of great importance to local companies, to enable them to be competitive with mainland companies, which in today’s globalised markets have become direct competitors.

It was also noted that in companies where the products have been around for a significant amount of time (several years), such companies are less willing to take up new products and try to achieve a larger market share. This is especially the case for subsdiary companies. In such cases, it is not unusual for the mother company to decide and impose the products that the company in question should manufacture, with these products potentially even being those that the mother company deems not worth marketing anymore, but is forced to continue producing for example for past use rs. Therefore, the common approach is for no investment to be considered in such
production lines. Further to this, some local companies were identified to be less willing to invest in new manufacturing approaches, with the mentality of “if we have managed during the past decade, we will manage during the coming decade as well”. Such a mentality exists especially in those companies that have very limited exposure to new and innovative ideas from outside the company.

During the discussions with manufacturing personnel, other factors not directly mentioned in the survey were also brought up. In more than one instance it appeared that the potential enlargement of the physical factory size was an issue. Companies seem to find it difficult to enlarge the physical size of the factories, with the main reason being attributed to the documentation and bureaucracy necessary for such a move. Another recurring issue in these discussions with several manufacturing organisations relates to the lack of networking between the local factories themselves especially in the case of micro and small companies. This is unfortunate for multiple reasons, including the consequent absence of sharing of potential solutions and of the identification of possible customers through subcontracting. Lack of networking also affects the already existing local automation service providers. Due to this lack of communication and networking, many of these service providers are still relatively obscure, with only a few being known by the majority of companies.

In concluding this preliminary investigation of the data obtained from the survey, one can note that for a survey of this nature, it is essential that a company tour be taken prior to the actual interview. The survey itself is at times too general and thus not capable of capturing all the information that is being obtained through the tour and the discussions being taken up with the respondent.

6. CONCLUSION AND FUTURE WORK

Further surveys will be performed with more companies from the various outlined manufacturing sub-sectors. This will give the possibility to analyse the various sub-sectors, together with the company size categories, on their own. Following this, the perceptions and problems related to manufacturing currently existing within particular sub-sectors could be determined. After an-depth analysis of the identified manufacturing perceptions and problems, the project will then proceed to address these specific manufacturing issues that emerge from this study.

ACKNOWLEDGEMENTS

This work is funded under the Maltese National Research and Innovation Programme through the Malta Council for Science and Technology. The research team would also like to thank all those companies that have till now participated in this study.

REFERENCES

A Research on Fault Detection and Classification of Cyclic Signals Using Spline Regression and Support Vector Machine

Jonghyuck Park, Sung-Shick Kim, and Jun-Geol Baek*

Graduate School of Information Management and Security
Korea University
Seoul, 136-713, Republic of Korea

ABSTRACT

Various fault detection methods based on statistical concepts such as non-stationary SPC chart, PCA, PLS, Hotelling’s T2 are widely used in order to detect the fault of the cyclic signal. In practice, even for the identical processes, however, the physical time lengths of cycles differ, and often signals are not available from some designated sampling points for various reasons, these are missing data. Therefore those classical methods are not suitable for real world process data which has different sampling lengths even missing values.

In this paper, a new fault detection method for stepwise cyclic signal using spline regression and support vector machine (SVM) is proposed. Regarding step changing point of data as knot point, a cyclic signal is regressed on the spline function. After then, features are extracted from the regressed spline function as the value of coefficients which is multiplied by each basis functions. With these features, it becomes much easier to attribute the reason of fault to actual trouble step. By using those features from the normal and faulty sets, SVM classifiers are constructed. Those constructed SVM classifiers are not just able to detect the relation between the normal and the fault but also able to classify between the faults. Furthermore, by using fault data as well as normal data, SVM classifier is expected to asymptote to optimal classifier. Performance evaluation of the proposed methodology is provided.

1. INTRODUCTION

Industrial revolution has greatly influenced on the fields of manufacturing. It enables mass production by using manufacturing machineries; changes the main cause of quality. Quality, which was effected by experiences of workers, nowadays, depends on the status of production equipment [1]. To improve quality, various methods to measure the manufacturing devices have been developed. Among them, Shewhart control chart, proposed by Walter A. Shewhart of Bell Laboratory in 1920s, is the most widely used and well developed method in theory [2].

As technology becomes more complex, manufacturing processes need more delicate method to control. Recently lots of manufacturing processes require, in general, controlling dynamically varying abundant variables under a closed-loop. Moreover, developments of manufacturing technology and information technology make it possible to use real-time data that reflects the status of manufacturing processes. Consequently, the concept of process control is changing to controlling non-stationary variables rather than keeping the stationary attribute value. Hence, it is important to have a process controller designed to maintain satisfactory operations by compensating for the effects of disturbance and changes occurring in the process [3].

Even though, there exist variations unable to be controlled by process controller. Fault is defined as a change, of parameters or characteristic properties which reflect the status of processes, which cannot be handled by controller [3]. If any variable exceeds the limit of acceptance, that situation can be regarded as a fault. By applying that philosophy, the statistically computed acceptance could be set as a criterion for detecting faults. Those procedures are called fault detection, determination of the faults present in a system and the time of detection [4].

To conduct fault detection for non-stationary signal, analysis methods have been proposed regarding a sampled point over given time horizon as a variable. Non-stationary SPC chart, which sets up the upper control limit (UCL) and lower control limit (LCL) by computed mean and variance of each time-point variable, is very popular method for detecting faults. Applying CUSUM chart [5] or EWMA chart [6] to time-point variables, in addition, has been proposed which enable to analyze in view of time series [3]. In case that signals have large length, then the number of chart for variables to control increases. Moreover, those control charts do not consider the inter-effects between the variables. Multivariate statistical analysis methods, such as Hotelling’s T2 chart, principal component analysis (PCA)

* Corresponding author: Tel.: (82) 2-3290-3396; Fax: (82) 2-929-5888; E-mail: jungeol@korea.ac.kr
Flexible Automation and Intelligent Manufacturing, FAIM2009, Teesside, UK

partial least squares (PLS), are adopted for this reason [7]. To enhance the ability of analyzing effects caused by time, researches on fault detection method using wavelet have been conducted [8, 9].

These methods have difficulties to deal with the real-world data. Cyclic signal, generated from general processes, is a sequence of steps which have common characteristics. Even for the identical processes, the physical time lengths of cycles differ as shown in figure 1. At the point of step-changing, due to high variances of signals, probability to commit type II error increases. Often signals are not available from some designated sampling points for various reasons, what are called missing data. In this case, estimating the status of the time has difficulties when sampled data are missing. Fault detection method to deal with real-world data needs pre-processing algorithm to adjust the physical length of signals so that each signal has same length.

![Figure 1: Heterogeneous lengths of steps [10]](image1)

However, there exist tendencies to lose the characteristics of signal and to magnify the effect of unexpected noise when pre-processing algorithms are applied. Figure 2 and 3 show factors that affect the value of sampled signal. In figure 2, \( f(x) \) means designated pattern of signal by a process engineer. Contrary to the engineer’s hope, it is natural that white noises \( \epsilon \) are added to the designed pattern. Process system, in accordance with a recipe originated by the process engineer, follows a pattern of \( f(x) \) combined with \( \epsilon \).

![Figure 2: Nature of signals – in continuous field [10]](image2)
A Research on Fault Detection and Classification of Cyclic Signals Using Spline Regression and Support Vector Machine

Figure 3: Nature of signals – in discrete field [10]

It is impossible to know what are happening in continuous fields. Data we have are signals sampled at a specified interval of time. Furthermore, measurement errors that occur in sensor devices, $e_t$, are also added. Signal sampled at time $t$, $s_t$, can be expressed as the summation form of these factors $s_t = f(x_t) + e_t + e_t$. Thus, ultimate objective of fault detection is examining the variations in continuous fields rather than that of gathered signals [10].

In this research, feature extraction method using spline regression is proposed which prevents trends of losing characteristics of signals and magnifying undesirable noise. Following to [10], a problem exists in scale. In this time, features are divided by its variability calculated as standard deviation. By doing this, features are stabilized. Fault detection algorithm conducted by support vector machine classifier is introduced in [10] constructed on the features extracted by proposed method. In addition, support vector machine fault classification method will be discussed. Benchmark test is given to analyze its performance and remarkableness in comparison with conventional method.

2. **SPLINE FEATURE EXTRACTION**

Spline is a function defined piecewise by polynomials on partitioned domain. In this case, by focusing on the univariate polynomial spline, spine function $S(\cdot)$ defined on $[a, b]$ maps them to $\mathbb{R}$, the set of real numbers. $S(\cdot)$ is composed of real-valued piecewise continuous functions $P_i(\cdot)$ defined on a subinterval $[t_i, t_{i+1})$. That is,

$$ S: [a, b] \rightarrow \mathbb{R} $$
$$ P_i: [t_i, t_{i+1}) \rightarrow \mathbb{R} $$

such that $a = t_0 < t_1 < \cdots < t_{k-2} < t_{k-1} = b$ and

$$ S(t) = \begin{cases} 
    P_0(t), & t_0 \leq t < t_1 \\
    P_i(t), & t_i \leq t < t_{i+1} \\
    \vdots \\
    P_{k-2}(t), & t_{k-2} \leq t < t_{k-1}.
\end{cases} $$

The given $k$ points $t_i$ are called knots [11]. Spline functions are classified, based on the way it is defined, into cubic spline, natural spline, smoothing spline and so forth.

General method to represent a form of spline is linear basis expansion. For a $p$ dimensional vector $x$, transformation function $h_m(x)$ is defined on $\mathbb{R}^p$ maps them to $\mathbb{R}$. By linear basis expansion, $f(x)$ can be represented as linear combination of transformation function as shown in (3).
Flexible Automation and Intelligent Manufacturing, FAIM2009, Teesside, UK

\[ f(x) = \sum_{m=1}^{M} \beta_m h_m(x). \] (3)

\( h_m(x) \) called a basis of function \( f(x) \). Linear method can be easily adopted if every \( x \) in the domain is transformed through basis function. So that, the parameters \( \beta_m \), which shape the function, can be simply calculated. By generating basis functions determined by the type of spline and estimating the parameters \( \beta_m \) with linear techniques, spline function \( S(x) \) is acquired [12]. From now on, regard \( x \) as univariate variable which denotes sampling time of a signal; focus on cubic spline.

When cubic spline function has 2 knots, \( \xi_1 \) and \( \xi_2 \), the basis of the function can be defined:

\[
\begin{align*}
h_1(x) &= 1 & h_2(x) &= x \\
h_3(x) &= x^2 & h_4(x) &= x^3 \\
h_5(x) &= (x - \xi_1)^3, \\
h_6(x) &= (x - \xi_2)^3
\end{align*}
\] (4)

where \( \eta \) is a function that has a value \( x \) when \( x \) is greater than 0; 0 otherwise. This definition implies that \( h_5 \) and \( h_6 \) have meaningful information only when \( x \) is greater than \( \xi_1 \) and \( \xi_2 \) respectively. These characteristics are different from those of \( h_1, ..., h_4 \). So, we can regard the parameters \( \beta_1, ..., \beta_4 \) multiplied to \( h_1, ..., h_4 \) as features that reflect whole trends of a signal. In a different way, \( \beta_5 \) and \( \beta_6 \) multiplied to \( h_5 \) and \( h_6 \), which have meaning only on fixed region, can be considered as feature of specified regions. Also cubic spline function is second-order differentiable, \( \beta_5 \) and \( \beta_6 \) are able to be interpreted as evaluations of variation between the subintervals.

With these characteristics of spline, feature extraction for cyclic signals composed of some steps can be done efficiently. Data of cyclic signals we have is \( s_i \) sampled signal at a point of time \( t \). By setting step changing point to knot point, sampled values are regressed on basis generated by cubic spline. If there are \( k \) steps, \( k+3 \) bases can be generated. Through the procedure, parameters \( \beta_1 \) to \( \beta_{k+3} \) multiplied to each basis are obtained; those \( k+3 \) values are treated as the features of cyclic signal \( s_i \).

Spline feature extraction capacititates to project cyclic signals, of which length are different, to same \( \mathbb{R} \) space, so that various statistical method can be easily applied. Since \( \beta_1, ..., \beta_{k+3} \) are the features that reflect the characteristics of a specified subinterval, it is possible to trace when faults occur. Unlike conventional fault detection methods, condition for homogeneous interval length and no missing points are not required by the reason that the method is based on regression which doesn’t require any strict conditions of missing data and same interval [10].

As stated in [10], however, these features are not in the same scale. Though the features well reflect the characteristic of each subinterval, the each variability of features is quietly different. So, these features need to be standardized. In this research, method to standardize via dividing by its standard variance is provided. Also, mean squared error of the signal is regarded as feature. Hence, a 5-stoped signal has 9 features.

3. Fault Detection and Classification Based on Support Vector Machine

Support vector machine is a method of supervised learning, which maximize geometric margin [13, 14]. Suppose a given set enables to supervise training; the set can be stated as (5).

\[ D = \{ (x_i, c_i) \mid x_i \in \mathbb{R}^p, c_i \in \{-1, +1\} \}_i \] (5)

\( x_i \), defined on \( p \)-dimensional space, is called input variable and \( c_i \) is called class label which corresponds to \( x_i \). Supervised learning is applicable due to class label \( c_i \) is known. Support vector machine classifier has a form of hyperplane \( w \cdot x - b = 0 \) which separates two classes. Geometric margin \( 2/\|w\| \), which represented in Figure 4, should be maximized. Without loss of generality, maximize \( 2/\|w\| \) problem can be reformulated as minimize \( \|w\|^2 \).
In order to construct support vector machine classifier for inseparable case, shown in Figure 4, slack variable $\xi_i$, which represents degree of misclassification is introduced. Penalty cost $C$ multiplied by the summation of all $\xi_i$ is added to objective function which minimizes the inverse of squared margin. From this, the optimization problem (6) can be formulated, that is,

$$\begin{align*}
\text{(P)} \quad \min & \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad c_i (w \cdot x_i - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n
\end{align*}$$

and its dual program is obtained[12] as (7),

$$\begin{align*}
\text{(D)} \quad \max & \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j c_i c_j (x_i^T x_j) \\
\text{s.t.} & \quad 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, n \\
& \quad \sum_{i=1}^{n} \alpha_i c_i = 0.
\end{align*}$$

As a result of solving (7), support vector machine classifier can be defined as (8), where $\text{sgn}(x)$ is a function that returns $+1$ when $x$ is greater than 0; $-1$ otherwise.

$$f(x) = \text{sgn}(w^T x + b) = \text{sgn}\left(\sum_{i=1}^{n} c_i \alpha_i (x_i^T x) + b\right)$$

The greatest advantage of support vector machine, originally linear classifying method, however, is availability that can be transformed to nonlinear classifying method by using kernel trick. Only positive definiteness of kernel function is satisfied, support vector machine can be altered in numerous different shapes. For any positive definite kernel $K(x, y)$, dual problem and classifier can be formulated as (9).

$$\begin{align*}
\text{(D)} \quad \max & \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j c_i c_j K(x_i, x_j) \\
\text{s.t.} & \quad 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, n \\
& \quad \sum_{i=1}^{n} \alpha_i c_i = 0
\end{align*}$$

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} c_i \alpha_i K(x_i, x) + b\right)$$
By applying various kernel functions and adjusting its shape parameters, more accurate and asymptotic classification is feasible to be done.

In former research [10], extracted features through using spline regression are used as the set of support vector machine fault detection. Just data which is known as fault or normal exist, training set can be formed; support vector machine classifier is constructed. Continuing the former research, in this paper, method to classify the types of fault is presented. At first, fault detection is performed on the given query data. In case that the query data is decided as fault, classifiers constructed to divide the classes of fault are at work to determine the fault-class of the data. Suppose there are three classes of faults, fault-class A, B, C, three classifiers should exist to separate between A and B, B and C, C and A. By majority voting from the results of three classifications, fault-class of query data is obtained.

4. NUMERICAL EXPERIMENT

Using real data originated from semiconductor/display FAB process is most desirable to evaluate the performance of fault detection algorithm for the case that it applied to real case. Due to security policy, evaluation results of real data cannot be written in this paper. Accordingly, artificial data are used which resembles chiller temperature in photolithography process.

4.1. DESCRIPTION OF DATA

As shown in Table 1, a single signal is composed of 5 main steps. Each step has sampled length as represented in Table 1; its length can vary within the range of ±5%. During the main step, signal will follow the defined function. White noises, however, which follow normal distribution with mean 0 and standard deviation 3, are added to the whole process signal. Figure 4 shows the prototype of generated signal.

![Figure 4: A prototype of generated signal](image)

Table 1: \( f(x) \) of generated signal

<table>
<thead>
<tr>
<th>Step</th>
<th>Avg. Length</th>
<th>Length Diff.</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>± 5%</td>
<td>15(\sin(x\pi/80))</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>± 5%</td>
<td>15[\sin((x-40)\pi/80 + 3\pi/2) + 2]</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>± 5%</td>
<td>15[\sin((x-40)\pi/80 + 3\pi/2) + 2]</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>± 5%</td>
<td>15[\sin((x-40)\pi/80 + 3\pi/2) + 2]</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>± 5%</td>
<td>15[\sin((x-40)\pi/80 + 3\pi/2) + 2]</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The types of faults

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>type 1</td>
<td>Mean shift (+4) in step 2</td>
</tr>
<tr>
<td>type 2</td>
<td>High variation (std = 5) in step 3</td>
</tr>
<tr>
<td>type 3</td>
<td>Irregular step length (&gt;5) in step 4</td>
</tr>
</tbody>
</table>
Three types of faults are defined (Table 2). Type 1 faults are defined as occurrence of mean shift with amount of +4 in step 2. If a signal is classified as type 2 faults, that signal has high variance white noise in step 3. Finally, a signal is regarded as type 3 faults when lengths in step 4 are distinguishably long.

### 4.2. DESIGN OF EXPERIMENT

The sets of randomly generated signals are formed to evaluate the performance of methods. Three sets are generated. First set is composed of normal signals in the proportion of 70% and rest of signals follow type 1 fault. In the same proportion, set 2 and 3 are the mixture of normal signals and type 2-3 fault ones respectively. Each set consists of 100 signals.

Not only verification of fault detection sensitivity, but also validation for fault classification should be done. For total 200 signals, 70% of them are regarded as normal, rest of them are split into three types appeared in Table 2 with same ratio. The fault detection and classification are done for the generated set following to the methodology introduced in chapter 3.

For evaluating generated data sets, performance test was done. These results from the 5-folds cross validation test. In order to compare fault detection algorithms, total 7 methods are considered. Non-stationary chart is one of them. Following to convention, 3-sigma control limits are set up. However, tolerance limit for faulty point is fixed for the sake of coordinating tradeoff between type I error and type II error. For extracted features by cubic spline method, Hotelling’s \( T^2 \) method and support vector machine are considered. The probability of Hotelling’s \( T^2 \) defines the size of ellipsoid, is set as 0.5, 0.7 and 0.99 to modify its ability to detect faults. In the case of support vector machine, Gaussian kernel and inverse multi-quadratic kernel with are considered as the control part of this research.

### 4.3. RESULT

In those tables, 'alpha' stands for the probability of occurrence of type I error - regarding as fault for normal, and 'beta' stands for the chance of type II error – vice versa. As shown in Table 3, overall performances of fault detection methods which use extracted feature are quietly better than non-stationary SPC chart in type 1; however, in the case of fault type 2 and 3, feature based fault detection methods work worse than non-stationary SPC chart. It is mainly affected by the characteristic of estimator. Features, generated by ordinary least squares method, are based on the assumption of homoskedacity. Even the mean squared error is regarded as feature, homoskedacity assumption is so strong, and that leads to high beta. However support vector machine is better than \( T^2 \) method in detecting faults. Unlike other methods, support vector machine uses faulty data for training. Hence, asymptotic fault detection is expected to be done when fault detection is conducted via support vector machine.

**Table 3: The benchmark result for the detection algorithm**

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alpha</td>
<td>beta</td>
<td>alpha</td>
</tr>
<tr>
<td>SPC Chart</td>
<td>.000</td>
<td>.533</td>
<td>.000</td>
</tr>
<tr>
<td>sigma = 3 Tolerance 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC Chart</td>
<td>.200</td>
<td>.067</td>
<td>.200</td>
</tr>
<tr>
<td>sigma = 3 Tolerance 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T^2 ) with</td>
<td>.071</td>
<td>.000</td>
<td>.071</td>
</tr>
<tr>
<td>probability 0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T^2 ) with</td>
<td>.029</td>
<td>.000</td>
<td>.029</td>
</tr>
<tr>
<td>probability 0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T^2 ) with</td>
<td>.014</td>
<td>.033</td>
<td>.014</td>
</tr>
<tr>
<td>probability 0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian Kernel</td>
<td>.029</td>
<td>.067</td>
<td>.214</td>
</tr>
<tr>
<td>Inverse Multi-Quadratic</td>
<td>.029</td>
<td>.033</td>
<td>.214</td>
</tr>
</tbody>
</table>

Classification performance, shown in Table 4, is computed by this procedure. After fault detection done by \( T^2 \) method, if a signal is regarded as fault, classification is done by sequential use of support vector machine classifier. Its performance is well enough because support vector machine is able to determine tiny differences structurally. By suitable use of proper method, fault detection and classification performance could increase significantly.
Table 4: The performance of the proposed classification algorithm

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Classification Method</th>
<th>Detection</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 with probability 0.95</td>
<td>Gaussian Kernel</td>
<td>.029</td>
<td>.389</td>
</tr>
<tr>
<td>T2 with probability 0.95</td>
<td>Inverse Multi-Quadratic</td>
<td>.029</td>
<td>.389</td>
</tr>
</tbody>
</table>

5. Conclusion

In this study, to detect the faults of cyclic signal, cubic spline regression and support vector machine fault detection and classification are presented. However, there are something remained to be properly used. First, feature selection algorithm should be enhanced. As shown in Table 3, feature selection can detect fault which has mean shift; other faults are not able to be detected easily. Assumption on homoskedacity will be the main reason of poor result of fault with high variance white noise. A feature that reflects the variance of a step should be proposed to detect faults like type 2. Also, step length should be considered as feature of a fault. Those suggestions enlarges feature dimension, however, it is expected to detect faults more sensitively.

In order to enhance the performance of support vector machine fault detection and classification, kernel selection method should be considered. Like selecting the probability in T2 method, parameters that shape kernel should be selected. However, unlike T2 method, there is no explicit relation between parameter and type I-II error. Also the parameter as penalty cost for non-separable cases should be chosen. More knowledge on the relation between support vector machine and the characteristic of signal is hoped to enhance this proposed method.

References

The project of a visual programming interface for the Mitsubishi Movemaster RV-M1 robot

Krzysztof Foit*, Daniel Reclik

Institute of Engineering Processes Automation and Integrated Manufacturing Systems
The Silesian University of Technology
Konarskiego 18A St.
Gliwice, 44-100, Poland

ABSTRACT
The Mitsubishi Movemaster RV-M1 is a small industrial robot with a five-axis manipulator and maximum payload of 1.2 kg. The main advantage of this robot is that it could be simply programmed using any personal computer equipped with the RS232 or the Centronics port. An operator writes a program in Movemaster Commands (MRL) language, then uploads it to the robot's controller. The controller memory is divided into two regions: one for commands and the other for positions. The positions could be programmed, verified and modified using a teaching box, but the program cannot be written or altered in this way – it is needed to use a computer. Most of available software, including the one created within the framework of scientific projects, is offered with a program editor, simulation window and even some diagnostic tools. The programming process is conducted in a standard way by writing commands and parameters. There is a lack of visual programming or creators (often called wizards) to carry on typical tasks. The aim of this paper is to present a project of the new software, which could simplify the programming process by introducing "wizards", templates for commonly used tasks and some elements of visual programming. Because the software is in the early development stage, some MRL commands have not been implemented yet.

1. INTRODUCTION
The introduction of visual programming has been a milestone in the field of application development. Manipulating elements graphically have been easier than specifying it in a textual form. On the other hand, this manner of the program description is often criticised because complicated algorithms can not be displayed in a clear form like text listing, however graphical representation is often used in systems’ modelling. This paper presents a project of a simple visual programming system intended for use with the Mitsubishi Movemaster RV-M1 robot.

The Mitsubishi Movemaster RV-M1 robot has a five-axis manipulator and is equipped with a compact control unit, which could be easily connected to any personal computer, using RS232 or Centronics port. Although the robot is quite small (the manufacturer calls it “the micro-robot system”), it is a complete, full-equipped industrial device. The complete system consists of three parts: the manipulator, the control unit and the teaching box. A typical configuration is shown in Figure 1. A personal computer (shown in Figure 1) could be any system equipped with the RS232 or the Centronics port – the first type of the port is preferred due to bi-directional transmission (the Centronics connection allows one-way communication: from a computer to the robot), however the transfer speed could be max. 9600 bps.

The memory of the robot is divided into two areas: the one for a program and the second for positions – thus the positions and the program could be changed independently. Positions can be uploaded to the robot's memory from a computer or by teaching. Every position entry has its own identification number, given by a robot’s operator, which is used in a program code. The robot could be programmed in two manners: by using a personal computer or by using a pre-programmed EPROM chip – there is no possibility of writing or change a program using a teaching box [1]. The second method is intended for infrequently reprogramming.
2. PROGRAMMING THE ROBOT WITH USE OF A PERSONAL COMPUTER

The minimum environment, required for programming with use of a personal computer is:

- any PC computer equipped with at least one RS232 or Centronics port,
- the Disk Operating System with any text editor, optionally with a high level programming language like BASIC, Pascal or C – it could be BASIC in ROM (in case of very old computers),
- a text display and keyboard – no mouse is required.

Commands could be sent to the control unit of the robot in two formats:

- without preceding line numbers – causing the immediate execution of the command,
- with preceding line numbers – causing the store of the commands in the control unit’s memory.

A program for the robot is written using Movemaster Commands [1]. Each line has the following syntax:

\[
<\text{line number}> \quad <\text{command}> \quad [\quad <\text{parameters}>]\quad .
\]

\begin{verbatim}
10 MO 10
20 DW 10,0,0
30 HE 11
\end{verbatim}

Fig. 2. A sample program written using Movemaster Commands

The Movemaster Commands are two-letter strings (abbreviations of the full name of the function), however parameters are usually decimal numbers with a preceding “+” or “-” sign. It is also possible to enter hexadecimal numbers using “&” character (for example “&2A”). An example of a program is shown in Figure 2. A program is transferred to the robot’s controller in the text form without any compilation or translation (as is). A basic problem
The project of a visual programming interface for the Mitsubishi Movemaster RV-M1 robot
during writing a program is to remember all the commands abbreviations and parameters. It is quite easy to write and execute a simple program, but developing the optimal code is more difficult, due to the mentioned problem.

The simplest method of transferring a program to the controller is to use the `COPY` command. For example, it could be entered in following manner:

```
COPY PROGRAM.ROB COM1:
```

Where `PROGRAM.ROB` is an example name of a file containing a program code and `COM1:` is the name reserved for first serial port in the system. It should be mentioned that the serial port should be earlier configured using `MODE` command.

### 2.1. The Overview of the Existing Applications Designed for RV-M1 Programming

The Movemaster RV-M1 is a product from the era when graphical interfaces were not commonly used. It has been gained popularity as a didactic machine due to its small dimensions and simplicity, but – being fully-equipped industrial machine – the robot has been also used in many scientific projects and works where high precision is required (like for example laboratory samples preparation) [2-6]. Most of the existing software has been either intended for didactic purposes or for simulation. The most known software is the COSIROP and COSIMIR packages [7]. They are commercial products with most advanced tools for programming, monitoring and simulation of the robot. The other one is Denford VR Robot intended mainly for simulation – this is also commercial software. There is also the software licensed under BSD License called JRobot [8]. It is the Java-based package developed for simulation, programming of the RV-M1 robot that also allows a remote operation and programming through a server connected directly to the robot.

A similar idea as for JRobot has been the beginning of a development of a software package at the Institute of Engineering Processes Automation and Integrated Manufacturing Systems [9,10]. It has been designed for use with the RV-M1 robot and consists of three applications: ROBO (used for programming and simulation in a virtual environment), SERWER (a network server application used for joining ROBO’s virtual environments) and KAMROB (an application which allows connecting a real robot with the virtual one in ROBO application). The main advantages of the package are:

- programming and simulation of the RV-M1 robot in a virtual environment,
- the possibility of joining virtual environments of the ROBO application working on different machines over the network and create a common workplace,
- programming the real RV-M1 and executing the commands remotely (over the network),
- remote supervision of the workplace of the real robot through an Internet camera,
- the possibility of generation of a simple program using a PTP trajectory generated in a CAD application (including a gripper status: open or close).

The software package has been developed in co-operation with the student within the framework of first and second semestral work and MSc diploma thesis. In the Figure 3, the main window of the ROBO application is shown. The aim of the package creation has been to get the software able to control a robot via Internet connection (so called “telerobot” software), using an advanced user interface and simulation tools. The RV-M1 has been an ideal machine to achieve a proper level of safety – small, lightweight and its workspace could be easily separated from the other parts of a laboratory. Due to modular structure, the software could be adapted to work with other type of robot. In the elaborated package, some elements of non-conventional programming have been introduced: the program generation on the base of a trajectory drawn in CAD software. Initially, this feature was a standalone application, now it is included in the package. It gives some advantages by generating the skeleton of the program automatically, but the advanced commands still must be entered manually.

It should be remarked that there is no software intended for the RV-M1 robot that significantly facilitates the programming process. In general, a program must be written by entering commands and parameters in a text form. It requires that the operator should be familiar with programming of the particular system. A visual programming application could improve this process in case of being an average user and could be a good tool for a novice user.
3. THE PROJECT OF A VISUAL PROGRAMMING SOFTWARE FOR THE RV-M1 ROBOT

In this section, a general concept of a visual programming environment will be presented. Furthermore, the first test version of this software will be shortly described.

The main assumption during creation of an overall conception of the software has been the simplicity of use. The second important thing has been a possibility of running the software on older computers because there is not necessary to engage a high processing power. On the other hand, newer machines do not often have the RS232 port – it is possible to use USB-serial converters, but sometimes it could cause some software problems. To achieve more flexibility and possibility of adapting the application to the other system than RV-M1, it has been decided that definitions of commands and its correspondent abbreviations will be placed in a dictionary file, compatible with any text editor.

Because of a simple syntax of the Movemaster Commands language, it has been decided that the first version of the software will not use a typical visual programming model (“boxes and arrows”), but it may be introduced in future versions. Some advanced functions have not been implemented yet. The application contains only the most used functions – mainly connected with positions, movements and the gripper. Moreover, the following assumptions have been made:

- all the functions have been divided into six groups, as it has been done in user’s manual: position/motion instructions, program control instructions, hand (gripper) control instructions, I/O control instructions, RS232 Read instructions and miscellaneous instructions,
- every function has been represented by its full name,
- every parameter has been described by its full name,
- a user interface should be as intuitive as possible,
- the process of a program creation should be done mainly by using a pointing device,
- there should be no need to do any post processing actions – it could be able to transfer a program to the robot directly from the application.

The main window of the created application is shown in Figure 4. It consists of a pulldown menu section, commands tree and two fields: one for a program and second for commands’ parameters. The pulldown menu has the standard File section, where typical commands like New, Open and Save are placed. The Transfer section contains
commands used for communication with robot: it is Upload and Download for transferring a program to and from the robot, Clear for erase the program and/or positions stored in the controller, Run for a program execution and RS232 Config for setting the basic parameters of the serial port.

![Image of visual programming application](image.png)

Fig. 4. The window of the visual programming application for the RV-M1 robot

The process of programming of the robot with use of the visual programming is conducted in the following manner:

- if it is necessary, the controller’s RAM memory could be cleared by selecting Transfer -> Clear from the pulldown menu,
- first of all it is needed to define positions – it could be done using the teaching box interface,
- next, the program could be created – it could be done by selecting a proper instruction from the tree and putting it in the program list (the upper one) by drag&drop or double click,
- the added instruction has been added on the list and marked for further edition; at the same time the parameters of that instruction will be shown on the bottom list – the values should be entered by selecting the parameter and entering values,
- after entering the last instruction the program could be saved on disk, then transferred to the robot by selecting Transfer -> Upload from the pulldown menu,
- after the upload, the program could be executed by selecting Transfer -> Run from the pulldown menu.

As it has been presented, the programming process is quite simple. The main advantage of this method is that full names of commands and parameters are used instead of incomprehensible abbreviations.
4. SUMMARY

In the paper, the project of visual programming interface for the Mitsubishi Movemaster RV-M1 has been presented. The developed application is still at the test stage and some instructions of the Movemaster Commands are not implemented yet. This concrete version of the software is intended mainly for educational purposes and it could be used during laboratory classes in order to collect users’ opinions. It is planned to introduce a package of “wizards” and templates for the most common tasks. In order to keep the flexibility of the application “wizards” should be implemented in some type of macro language. The future work will be focused on development of a better interface and functionality improvement.

REFERENCES

Automated Deburring System with a 3-Axis Robot for Injection Molded Parts

Seong-Jin Kim¹, Dong-Won Kim²*, Gyun-Eui Yang³, Eun-Young Heo², Chan-Gie Lee², and F. Frank Chen⁴

¹Automobile Parts & Mold Technology Innovation Center, Palbokdong2ga, Duckjingu, Jeonju, 561-855, South Korea
²Department of Industrial & Information Systems Engineering and The Research Center of Industrial Technology, Chonbuk National University, Jeonju, 561-756, South Korea
³Department of Mechanical Engineering, Chonbuk National University, Jeonju, 561-756, South Korea
⁴Department of Mechanical Engineering and Center for Advanced Manufacturing & Lean Systems, University of Texas at San Antonio, TX 78249, USA

ABSTRACT

Demand for plastic molded parts is on the increase in the automobile industry, since they can reduce the costs and total weight of an assembled car without lowering its structural performance. However, burrs are commonly found on the exterior of injection molded parts and these need to be removed. The burrs can be removed manually but, the manual operations are very labor intensive, require highly skilled staff and can be hazardous due to the dust that is generated. This paper presents a study on an automatic deburring system for injection molded, automobile parts, by using a 3-axis robot arm. The system is composed of a robot arm for removing burrs, a pallet and a conveyer for transferring the parts, a pneumatic clamp for fixing of the parts, and a programmable logic controller for performing the overall system control. An example is shown for injection molded parts of an automobile to demonstrate the effectiveness of the developed system.

1. INTRODUCTION

Recently, demand for plastic molded parts has greatly increased in the automobile industry because they can reduce the total weight of an assembled car without lowering its structural performance. Furthermore, the plastic molded parts can be mass-produced at lower cost, compared to other machined or forged parts. However, burrs are commonly found in injection molded parts, on their exterior, like machined ones. Note that burrs are brought about by the machining processes producing plastic deformed shapes, but they are commonly found in plastic molded parts as well. The burrs on the molded parts occur due to various causes such as high injection pressure, low bearing forces of molds, high flow rate of plastic resins, and surplus resins against cavity capacity, etc. They usually have broken, or sharp edges that could damage the contact faces of the assembled parts, as well as affecting the allowed tolerances. The burrs need to be removed while still on the shop floor and conventional manual deburring/polishing operations have been used. However, these manual operations are very labor intensive, require highly skilled staff, can be error-prone after long periods of continued duty and can even be hazardous due to the dust that is generated which consequently decreases the productivity of the shop floor.

Much research has been conducted in the development of automated deburring systems for dies and molds, which use robots and/or other transfer machines. Some researchers have focused on the prevention of burr formation: Ko [2], and Chang [3] tried to propose optimal tool paths thus avoiding the formation of burrs as well as predicting the burr
creation processes. Kim and Ko [4] presented a method of measuring burr shapes and Ko et al. [5] developed an expert system to predict burr shapes in milling operations. However, most researchers have considered the development of efficient burr removal systems, by robots or other NC systems, ultrasonic devices, electrolyzing methods, etc. Shin et al. [6], Yun et al. [7], and Ryuh and Oh [8] employed robot systems for performing deburring tasks, with sensors such as vision. Choi [8] proposed a micro deburring technology through ultrasonic vibration. Lee et al. [9] also introduced a micro-level deburring technology for sheet metals by using electrolysis and Liao et al. [10] have proposed a dual-purpose compliant tool head for automated machines to ensure stable polishing/deburring force control. This study reports the development of an economical, 3-axis robot deburring system dedicated to plastic molded parts, which is controlled through a programmable logic controller (PLC).

2. AUTOMATED DEBURRING SYSTEM

The developed robot deburring system is composed of a 3-axis, prismatic robot for burr removal, a pneumatic clamp for workpiece stability, a part transferring carriage installed on a guided rail, a conveyer for moving the parts, and finally, a PLC for controlling the entire device modules, as shown in Figure 1.

The burr removal process is as follows: A plastic molded, raw part comes to the deburring system through a part conveyer. Then, the part is moved onto a pallet by two cylinders in the carriage, being exactly aligned to a target position by a servo motor. The part is fixed with a pneumatic clamp, and the burrs of the part are removed by an end milling set in a spindle mounted on the 3-axis robot. Finally, the de-burred part is returned to the conveyer by the part transferring carriage.

![Figure 1: Block diagram of an automated robot deburring system](image)

2.1. PART CLAMPING DEVICE

Vices or jigs may cause unnecessary marks on the faces of plastic molded parts and hence a pneumatic, vacuum clamp is used for the fixing of the plastic molded parts in this system. The vacuum clamp is free from any suction marks, has a secure fast clamping operation, and maintains robust fixing without any shaking or waving during the deburring operation. Further, the clamp ensures the pre-established part location so as to reduce set-up times. Thus, the clamp has to be equipped with a jig that has the same exterior as the part, such as the cavity shape jig shown in Figure 2. Furthermore, the clamp should not interfere with the operations of the deburring robot.
2.2. **PART TRANSFERRING DEVICE**

To support consecutive burr removal operations, several pallets are required, which correspond to the supply buffers for the deburring robot. The raw parts, set in the pallets, are moved to a vacuum clamp through a transferring carriage that includes two cylinders for gripping and elevating, and a linear slider driven by a servo motor (see Figure 3). The gripping cylinder, which loads a part containing pallet, should not leave any marks on the exterior of the part, and be able to hold the part without any unnecessary, dangerous motion during the movements to the vacuum clamp. Then, the part can be moved to the clamp by the linear slider, which is a sort of servo-driven, prismatic robot.

2.3. **DEBURRING ROBOT**

The main module of the automated deburring system is composed of a 3-axis prismatic robot and a spindle holding a deburring tool, which takes away the burrs from the part fixed by a vacuum clamp. The spindle provides secure and sufficient torque and is able to withstand machining loads when removing the burrs by the installed tool, e.g. an end mill. The 3-axis robot accurately moves the tool to the deburring position through its three prismatic axes. At this point, the robot is controlled by a MAC (a multi axis controller) by Samsung Electronics. Thus, accurate and robust tool paths, as in a NC-machine, have to be determined in advance by a robot process planning module, which should be post-processed to the robot data such as the cutter location data of a NC machine. The robot tool paths should be free from the possibility of collision with other peripheral devices such as the vacuum clamp and the transferring carriage.
2.4. **ROBOT SYSTEM CONTROL**

As mentioned in section 2.1, the automated deburring system consists of three sub-modules: a vacuum clamp, a transferring linear carriage, and a deburring robot. These modules have to maintain close relationships with each other, and are controlled by a central monitoring system, i.e., a PLC. The whole system control can be represented by a ladder diagram, as shown in Figure 5.

A prototype robot deburring system, developed in this study, is shown in Figure 6, and its major specifications are summarized in Table 1.
Automatic Deburring System with a 3-Axis Robot for Injection Molded Parts

(a) Transferring carriage    (b) Deburring robot    (c) PLC panel

Figure 6: A prototype robot deburring system

Table 1: Major specifications of the prototype deburring system

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strokes</td>
<td>200×200×100mm</td>
<td>Spindle: Hz</td>
<td>400</td>
</tr>
<tr>
<td>Maximum speed</td>
<td>1,000mm/s</td>
<td>Spindle: Volt</td>
<td>220</td>
</tr>
<tr>
<td>Repeatability</td>
<td>±0.02mm</td>
<td>Clamp size</td>
<td>80mm×80mm</td>
</tr>
<tr>
<td>Spindle RPM</td>
<td>10,000</td>
<td>Clamp suction</td>
<td>Above 5bar</td>
</tr>
</tbody>
</table>

3. A CASE EXAMPLE

In order to verify the performance of the developed prototype system, a fuel filler door from an automobile has been used as the plastic molded part for the deburring test. The part is made of ABS resin, being used for the fuel supply cap. The molded part has a burr of 1 mm thickness around its edges, as shown in Figure 7(b).

(a) A fuel filler door: finished part       (b) Burr in the edge of the part

Figure 7: Fuel filler door example

To obtain the cutter location (CL) data for the molded part, the tool path of the deburring robot is determined first. The fuel filler door has been modeled by CATIA V5.0 and 430 boundary points have been extracted from the edge of the door (see Figure 8), which is used for the CL data for the deburring tool set in the spindle of the robot system. To verify the deburring tolerance and the fixing operation of the vacuum clamp, a test operation was performed on the 3-axis machining center in advance (see Figure 9). As a deburring tool, a reverse R-end mill is used instead of a conventional formed end-mill that could cause adhesion to the cutter edge due to the heat occurring during the deburring process.
Throughout the experimental test operation, the proposed system has shown successful deburring performance and the cycle time of the developed system takes 30 seconds per part, which is an improvement of 15% in productivity over existing manual grinding operations. Moreover, from the aspect of system stability and aestheticism, the molded part is stably fixed with the vacuum pump and the machined surface is clear from the burr. Also, the developed system is able to support flexible management against changes in production as well as ensure applicability to other plastic molded parts.

(a) Modeling
(b) Point data
(c) Deburring data

Figure 8: Deburring data generation

(a) Part fixing by a vacuum clamp
(b) Deburring verification

Figure 9: Test deburring operation using a machining center

(a) Reverse R-end mill
(b) Sectional shape of the part and tool

Figure 10: Tool and its shape
4. CONCLUSION

Plastic molded parts are increasingly used as components in manufacturing automobiles since they can reduce the total weight of an assembled car without lowering the car’s structural performance. Also, they can be easily modeled by conventional commercial CAD/CAM systems but burrs on the face of the molded parts are commonly formed during the plastic injection molding processes, because of the injection molding conditions and the defects of the dies that are employed. In most cases, burrs are removed through manual grinding operations that decrease the productivity on the shop floors. Furthermore, the manual operations can often result in musculoskeletal disease of workers.

For the automatic deburring of injection molded parts, a prototype system has been developed in this study, which comprises a 3-axis prismatic deburring robot, a transferring carriage, a vacuum clamp, and a PLC for performing the overall system control. The transferring, clamping, and deburring operations have been verified through a test operation carried out on a 3-axis machine center, and the test results have shown successful performance with economical devices and tools, compared to existing 6-axis based, deburring robot systems, or other dedicated or compliant tool head based commercial equipment [11]. In particular, the developed system, presently used in a local company, has shown a 15% improvement in productivity over existing, manual grinding operations for the tested part. Further development underway includes the integration of the proposed system with an automatic, exterior data inspection system.

ACKNOWLEDGEMENTS

This study was a part of The Shop Floor Ordered Technology Development Program supported by Kunsan Innovational Cluster Propulsion Center, and the authors thank a subcontract factory S of GM Daewoo Auto & Technology Co.

REFERENCES

Obstacle and Collision Avoidance in Unstructured and Time-Varying Environments: A Screw Theory Approach for Manipulators

Carlos R Rocha*, Henrique Simas, Daniel Martins, and Altamir Dias

Department of Mechanical Engineering
Universidade Federal de Santa Catarina
Florianópolis, SC, 88040-900, Brazil

ABSTRACT

Robots are increasingly being used in applications where they are required to operate in unstructured or poorly mapped environments. Eventually, these environments can change over time, as in case of objects or people entering the robot’s task space. In many cases, human and robot safety issues are of concern, since there will be human-machine interaction. In those, obstacle and collision avoidance are of major concern, to prevent injuries to people and possible damages to the equipment involved. Introducing kinematic redundancy in the mechanical structure of the manipulator increases dexterity, allowing it to execute tasks while complying with complementary objectives, such as avoiding singularity, operating within joints/actuators limits and obstacle/collision avoidance. However, redundancy turns the resolution of inverse kinematic into an infinite solutions problem. This work presents a solution to the problem above, based on the screw theory, the Kirchhoff-Davies method to solve differential kinematics in closed-loop kinematic chains, and the use of virtual chains to define the task to be executed and to model the obstacles to be avoided. It will be presented an example of the use of these techniques to solve the inverse kinematics problem involving redundancy to achieve obstacle avoidance. Finally, the extension of the technique to avoid collision with moving obstacles or unstructured environments will be discussed.

1. INTRODUCTION

Industrial manipulators usually operate in structured and isolated workspaces, in order to avoid collision, which could damage materials and equipment, and also cause injuries to humans. This reality is gradually changing, as manipulators are increasingly being used in new applications, which are not only restricted to the usual safe workspace. In these applications, the environment is not necessarily known a priori, and it may be time-varying, with obstacles moving through the workspace. It could even interact directly with human beings, eventually. The ability to autonomously operate in this kind of environment is a key feature to intelligent robotic systems [1]. Several examples of these applications are presented in literature. Among others, it could be noted the human/robot interaction, where a collision could severe injury a human, and safety issues are especially important [18].

To guarantee integrity of all elements present in the workspace, it is fundamental to have collision avoidance strategies, with obstacle detection/deviance while the manipulator executes the task assigned. In this work, obstacles are defined as entities placed in the workspace, which contact with the manipulator is undesirable. It is also considered that collision may occur in any link of the manipulator kinematic chain, and not only in the end-effector.

In a completely structured known environment, it is possible to map all collision possibilities, which can be monitored and modeled, being part of the system model. Cheung calls this approach The Piano Movers Problem [3].

However, in the most general case, the information about the environment is assumed to be incomplete, and environment itself cannot be considered static. In this situation, it is necessary to use sensors to gather environmental information during the task execution. These sensors must cover not only the possibility of collision with the end-effector, but also that the collision can occur in any link of the manipulator. Several authors discuss this problem,

* Corresponding author: Tel.: +55(53) 9971-3148; E-mail: cticarlo@gmail.com , cticarlo@furg.br
using different approaches and sensor types to solve it [1,3-5]. Lumelsky discuss the requisites and uses of a sensitive skin, and its importance to intelligent robots operating in non-structured environments [6].

Obstacle avoidance implies additional constraints to the task to be executed. In this case, it is necessary to have redundant degrees of freedom to accomplish the task [7-8]. In this work, redundancy means kinematic redundancy, defined as the difference between the degrees-of-control and the connectivity between two links of a kinematic chain by Martins and Piga-Carboni [9]. In a simpler way, redundancy occurs when the mobility of the kinematic chain is greater than the necessary in order to the manipulator execute a task. It can be noted that redundancy can depend on the type of the task to be executed [8].

Redundant manipulators have increased dexterity, allowing the addition of complementary goals to the primary task, such as avoiding obstacles, singularities and joints/actuators limits, and to optimize performance criteria. Having an infinite number of solutions, the inverse kinematics problem allows specifying additional constraints to represent the complementary goals cited. Several authors have proposed methods to solve the inverse kinematics problem in redundant manipulators, using pseudoinverse jacobian, task augmentation or task priority, to cite some techniques. A discussion of these more usual techniques is made in [8]. In this work, it will be used the kinematic constraints method, which is based on screw theory, the Kirchhoff-Davies method and the use of virtual chains, as presented in [10-12]. Among other features, this method does not operate with dimensional inconsistencies, and has the property of conservative movement [12].

In the next section, the fundamentals of the kinematic constraints method will be presented. Section 3 will present an example of the use of this method, in a static ambient and a known obstacle. The extension of the technique to an unstructured, time-varying environment will be discussed in section 4, where the use of sensors to identify obstacles, the strategies to select the proper kinematic model in each state that the robotic system can be while executing a task, and problems related will be addressed. Section 5 will present the conclusion.

2. BACKGROUND

The kinematic constraints method is based on screw theory. It uses virtual chains to impose movement to the manipulator, to monitor and to define complementary kinematic constraints as well. The Kirchhoff-Davies method defines differential kinematics constraints between the links of a closed kinematic chain. In order to obtain joint position values, it is necessary the integration of the inverse differential kinematics obtained by use of the Kirchhoff-Davies method, which results in drift, caused by the numerical method used. The kinematic chain then opens, resulting in errors. The addition of error virtual chains allows to control and to minimize drift. This section will present shortly these fundamentals. Fontan and Simas have detailed and exemplified explanation in their work [11-12].

2.1. Screw Theory

Screw theory is a tool for static and kinematic analysis of mechanisms. It was first formulated by Mozzi, in 1763, and it was systematized by Ball in 1900. Later, Hunt and Tsai applied this theory in static and kinematic analysis of mechanisms [10]. A screw is a geometric entity composed by a directed line and a pitch, which can represent rigid body movements, according to Chasles Theorem [13-14].

Consider a point P, moving from a position P_i to a position P_2. The movement can be described by a rotation around a directed line, followed by a translation parallel to that line, as shown in Figure 1. The magnitudes of rotation and translation are related by a pitch h. Based on this representation, the movement can be represented by four parameters, which are an unitary vector parallel to the screw axis (s), a position vector to a point belonging to the screw axis (s_0), the value of the rotation around the screw axis (θ) and the magnitude of the translation parallel to the screw axis (r). With these four parameters it is possible to calculate a homogeneous transformation matrix [13]. As in Denavit-Hartenberg notation, it is possible to combine successive displacements applied to a rigid body, caused by the joint movements in a kinematic chain multiplying the corresponding homogeneous transformation matrices.

A twist represents the instantaneous movement of a rigid body relative to a referential system. It can be defined as \( \mathbf{s} = (\omega, \mathbf{v}) \), where \( \omega \) is the angular velocity of the body and \( \mathbf{v} \) is the linear velocity of a point in the body instantaneously coincident with the origin \( \mathbf{O} \). \( \mathbf{v} \) has a component normal to the twist axis and another component parallel to this axis. The twist can be represented by a normalized screw \( \mathbf{s} \) multiplied by a magnitude \( \dot{\gamma} \), as shown in Equation (1).
In a kinematic chain, a twist can represent the movement of a link relative to its predecessor. It is common, in robotics, the use of prismatic and rotational joints, which simplifies the normalized screw associated with the joints, since the pitch is equal to 0, when a joint is rotational, and it is equal to ∞ when a joint is purely translational. In Equation (2), $\mathbf{s}_r$ is a normalized screw for pure rotation, while in Equation (3) $\mathbf{s}_t$ is a screw for pure translation.

$$\mathbf{s} = \mathbf{S} \dot{\mathbf{q}} = \left[ \mathbf{s} \right]_{\mathbf{k}}$$

In Equation (4), $\mathbf{S}_l$ is the Jacobian mapping joint displacements to link movements in inertial coordinate system.

$$\mathbf{S}_l = \sum_{i=1}^{l} \left[ \mathbf{s}_i \right]_{\mathbf{k}} = \sum_{i=1}^{l} \mathbf{s}_i \dot{q}_i$$

Equation (4) can be rearranged to group normalized screws and joint variables, resulting in

$$\mathbf{S}_l = \left[ \mathbf{s}_1 \dot{q}_1 \quad \cdots \quad \mathbf{s}_l \dot{q}_l \right]^T = \mathbf{N} \mathbf{q}$$

2.2. THE KIRCHHOFF-DAVIES METHOD

Davies adapted Kirchhoff’s circulation law to the case of closed kinematic chains, in order to analyze its differential kinematics [10,12]. In his work, he stated that “the algebraic sum of the relative velocities of kinematic pairs in a closed chain is equal to 0” [15]. This adaptation allows to define a constraint equation to a closed kinematic chain, relating its joint velocities [12].
A simple example can be used to illustrate this method. Consider the four-bar linkage depicted in Figure 2. Joint A is actuated, while the others are passive. Since it is a planar linkage, the screw system order is three, having two linear components in $x$-$y$ direction and an angular component around the $z$ direction. The components of the normalized screw will be $N, \vec{P}$ and $\vec{Q}$ [12]. Using the Kirchhoff-Davies method, the sum of the joint velocities is zero, as shown in Equation (6), where $\mathbf{s}_i$ is the twist relative to joint $i$, and $\mathbf{0}$ is a column vector, with dimension equal to the screw system dimension.

$$\mathbf{s}_A + \mathbf{s}_B + \mathbf{s}_C + \mathbf{s}_D = \begin{bmatrix} \dot{s}_A \\ \dot{s}_B \\ \dot{s}_C \\ \dot{s}_D \end{bmatrix} \begin{bmatrix} \dot{q}_A \\ \dot{q}_B \\ \dot{q}_C \\ \dot{q}_D \end{bmatrix} = \mathbf{0}$$ (6)

Equation (6) can be rewritten as

$$\mathbf{N} \dot{\mathbf{q}} = \mathbf{0}$$ (7)

To solve this system and obtain the passive joints velocities, Equation (7) can be rearranged, partitioning the twists into primary (which magnitudes are known) and secondary ones (which magnitudes have to be obtained), as shown in Equation (8).

$$\begin{bmatrix} \mathbf{N}_p \\ \mathbf{N}_s \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{q}_p \\ \dot{q}_s \\ \mathbf{0} \end{bmatrix} = \mathbf{0}$$ (8)

The secondary joints values will be determined by Equation (9), if $\mathbf{N}_s$ is invertible. Otherwise, the system will be in a singularity [10].

$$\dot{q}_s = -\mathbf{N}_s^{-1} \mathbf{N}_p \dot{q}_p$$ (9)

Graphs are a very useful tool to analyze kinematic chains, specially closed ones, and it eases the Kirchhoff-Davies equations definition. Tsai explains in detail the use of graphs to kinematic analysis of mechanisms [13,16].

### 2.3. Virtual Chains

Campos defines a virtual chain as a kinematic chain used to observe and obtain information about real kinematic chains, and also used to introduce constraints/movements to the real chains [10]. A virtual chain must be: a) a serial chain; b) the twists representing joints displacements must be linearly independent; c) it can not modify the mobility of the real chain [11,12].

In planar case, virtual chains must have mobility equal to three, in order to comply with the characteristics [10]. Most common are PPR chains, composed by 2 links connected by a prismatic joint, with one end having a prismatic joint and another end having a rotational joint, and RPR chains, which differ from the latter by having rotational joints in both ends. PPR chains (Figure 3.a) can be associated to Cartesian coordinate systems, while RPR chains (Figure 3.b) can be associated to polar coordinate systems.

In spatial configurations, virtual chains must have mobility equal to six, and PPPS chains, which can be associated with Cartesian coordinate systems, are common (Figure 3.c).
In order to use the Kirchhoff-Davies method, a kinematic chain must be closed. In serial manipulators, virtual chains can be used to close the overall chain. Virtual chains can be also used to monitor certain movements, as the case of observing proximity of a link to an obstacle. Another use of virtual chains is the definition of constraints to the movement, which could be very useful in to solve inverse kinematics in redundant chains.

Figure 4.a depicts an example of closure of a serial chain of a planar manipulator, where a PPR virtual chain is used to impose end-effector position. Figure 4.b depicts a RPR, used to monitor the proximity of an obstacle. In spatial configurations, the use of virtual chains is analogous. Fontan presents several examples of the use of virtual chains, applied to serial and parallel mechanisms [10].

2.4. INVERSE KINEMATICS RESOLUTION

By using virtual chains to impose the movement, the Kirchhoff-Davies method can be used to solve the differential inverse kinematics, considering the virtual joints as primary and the real joints as secondary. Some care must be taken to choose secondary joints, in order to obtain an invertible $N_s$ matrix. With known joint velocities, it is possible to obtain the joint displacements, through integration of the joint velocities over time. Numerical methods are usually employed, since the analytical resolution is very complex in most cases [10-12].

Numerical methods generate drift, resulting in position errors. In closed chains, those errors will imply in chain opening. Simas propose one solution to this problem in [12], by addition of error virtual chains, to monitor drift effects and to minimize closure errors.

With the modifications, the new kinematic chain is described by Equation (10), where $N_e$ and $\dot{q}_e$ are added, corresponding to the error normalized screws and to the variation rates of error, respectively.

$$
\begin{bmatrix}
N_p & N_s & N_e \dot{q}_p & \dot{q}_s & \dot{q}_e
\end{bmatrix}^T = 0
$$

(10)
The velocities of the secondary joints are obtained by Equation (11), where $q_e$ corresponds to the positions of the error virtual chain and $K_e$ is a positive definite gain matrix.

$$\dot{q}_e = -N_s^{-1}N_p \dot{q}_p - N_s^{-1}N_e K_e q_e$$

(11)

Simas analyzes and exemplifies the procedure, proving the stability of this method [12].

3. Case Study

To illustrate the method, consider the RRRR planar manipulator shown in Figure 5. The second link may collide with an obstacle, and the proximity will be monitored. In this example, two virtual chains are used. The positional PPR virtual chain imposes the movement to the end-effector, while the obstacle RPR virtual chain is used initially to monitor the proximity between the second link and the obstacle. The primary joints, in the monitoring state, are $p_x$, $p_y$, $r_z$, and $D$, while the secondary joints are $A$, $B$, $C$, $r_z1$, $r_z2$, and $p_x$. $D$ was chosen arbitrarily, to define an invertible matrix $N_s$. The $p_r$ joint is used to monitor the obstacle. When its position value is below a pre-established threshold, the manipulator will change to collision avoidance state.

In this state, the kinematic model will be modified, in order to impose a constraint represented by joint $p_r$. This joint will be considered a primary one, while joint $D$ will be turned into a secondary one. Joint $D$ will be actuated, in order to achieve a configuration where the task can be performed while avoiding the obstacle.

![Figure 5: RRRR manipulator in presence of obstacles](image)

When the value of the value position of $p_r$ is over the threshold, the manipulator changes its state to monitoring, and the kinematic model turns back to the first specified. To simplify the example, the error virtual chains were not included, but they are essential to solve the inverse kinematics. A similar example is presented in detail in [11], where a PRRR planar manipulator is used, and a spatial case, with experimental results, is analyzed.

4. Extending the Technique to Use in Unstructured and Time-Varying Environments

According to Cheung and Lumelsky, a manipulator must have a trajectory planning system based on sensors, in order to operate in unstructured and/or time-varying environments [3]. It is necessary the ability to identify possible contact with obstacles with any part of the manipulator, and it is necessary to have the capacity to rapidly process large quantities of sensorial information, in order to achieve local modification of the global trajectory planned. This area of research is very active, with various approaches in the use of sensors and strategies of adaptation. In common, it can be noted the consideration of a kind of sensitive aura over the manipulator structure, inside which the space can be considered free of collision. As the manipulator moves, the aura moves with it. When some obstacle contacts this aura, its position relative to one or more sensors is determined, and a strategy to avoid collision is taken.

It can be considered that it is possible to detect the proximity of an obstacle, and its position relative to some part of the manipulator, the same way that monitoring using virtual chains. The virtual chain can be used only to constraint the movement in order to avoid collision, leaving the detection and monitoring to sensors.
There will be different states during the execution of a task, according to Fontan [11]. In the free movement state, one kinematic model will be used, where redundancy can be used to optimize some criteria, or to limit the actuation of some joint. In collision avoidance state, another kinematic model will be used, where virtual chains to constraint the movement will be used. Depending on the sensor, several collision avoidance strategies can be used, depending on the type of the obstacle identified. The IPA sensor presented in [5], for instance, can be used to map several points, which could be used to estimate the shape of the obstacle and to use a virtual chain more adequate to each case.

One problem with the model switching is the possibility of discontinuities in the movement, with undesirable effects over the joints and the actuators. Fontan and Santos discuss this problem, proposing some strategies to reduce the discontinuities during switching. This problem can be more relevant when the environment changes over time, because of the continuous and uncertain switching [11,17].

Another situation to be considered is how to handle the presence of multiple obstacles, and the strategies related. The example presented only one possibility of collision. In some cases, the same approach may be used with more obstacles, if it could be considered that only one obstacle is in proximity of collision at any time. But if various obstacles are in imminent collision, the strategy must be modified. Some authors developed strategies to handle this possibility, and it is necessary to determine how to handle that problem with virtual chains [3,5].

5. CONCLUSION

This work presented a technique used to avoid collisions with obstacles of a manipulator operation in a structured environment, and discussed the characteristics necessary to extent the technique to the case of an unstructured and time-varying environment. To avoid obstacles while executing the task assigned to it, the manipulator has to be redundant, and to solve its inverse kinematics the kinematic constraints method is used. It was presented the fundamentals to this method, which is based on screw theory, the Kirchoff-Davies method and virtual chains.

When executing a task where obstacles must be considered, the kinematic model of the robotic system may vary, depending on the operating status (needing to avoid an obstacle or freely operating), which demands a decision-making algorithm that will choose the appropriate kinematic model at each instant of the task execution.

It was discussed how to extend the technique to use in unstructured and time-varying environments, where sensor information is needed. The sensors must cover all surface of the manipulator, developing a kind of sensitive skin. When the sensitive aura contacts an obstacle, sensors provide information about the position of the obstacle relative to a link, and the kinematic model is modified, in order to consider the constraint needed to avoid collision with the obstacle. Different strategies can be implemented, depending on the obstacle type and the information obtained from the sensors. The proposed extension discussed will need further work, and results will be presented in future papers.

Finally, it is considered that another contribution of this work was the definition of obstacle and collision, in the context of the robotic manipulation tasks.

REFERENCES


An Experimental Analysis of Bernoulli End Effectors for Contact Manipulation of Irregularly Shaped Objects

Nebojsa I. Jaksic* and Gajendiran Mohandass

Engineering Department
Colorado State University - Pueblo
Pueblo, Colorado, 81001, USA

ABSTRACT

Bernoulli end effectors or Bernoulli pickups are manipulating devices that use the Bernoulli effect to perform contactless manipulation of flat parts. They are often used for handling silicon wafers during the manufacturing of integrated circuits. In general, a Bernoulli end effector consists of a flat plate with a through hole placed in the center of the plate and connected to a positive pressure fluid source. To operate, the plate is placed close to a flat part to be manipulated, and then fluid is allowed to flow out of the hole and into the space between the plate and the part. This fluid flow creates a low-pressure zone which in turn creates an attractive force between the part and the plate enabling levitation of the part near the plate.

In this experimental study, a number of tests are performed to optimize the lifting capacity of Bernoulli end effectors in contactless mode. The lower flat surface of the plate of a Bernoulli end effector is changed into a conical surface with the center at the center of the hole. An optimal cone angle, discovered experimentally, increases the lifting capacity of the end effector about 20%. Additionally, a novel application of Bernoulli end effectors dealing with manipulation of irregularly shaped objects in contact with the end effector is analyzed. A number of experiments are performed to illustrate the versatility of these devices. The same Bernoulli end effector is used in contact mode to manipulate plastic bottles in various orientations as well as rocks. An advantage of using Bernoulli end effectors in contact mode instead of in contactless mode is illustrated by tilting an object from 0 to 90°. This is not possible in contactless mode since the low friction allows the object to slide off the plate when tilted.

1. INTRODUCTION

The scope of traditional end effectors like mechanical grippers or vacuum pickups is limited by the size, geometry, weight, and other physical characteristics of the manipulated objects. Industrial manipulators equipped with specialized robotic grippers are well suited to pick-and-place tasks in applications where the physical characteristics of manipulated objects are similar [1]. However, a simple change in object size or material may require a change of the gripper. Also, handling large, flat, non-rigid, or delicate parts is not easily accomplished with many standard grippers. Vacuum pickups used as end effectors further extend the range of objects a robot can manipulate. However, they rely on vacuum created between one or more suction cups and the manipulated object. If a sufficient vacuum cannot be established between the vacuum pickup and the object, then the object can not be manipulated. This is the case with porous objects or irregularly shaped objects where the objects’ surface characteristics do not allow creation of vacuum between the suction cups and the objects.

A Bernoulli end effector or a Bernoulli pickup is a pressurized fluid device capable of picking up objects by creating a difference in pressure between the surface of an object and a surrounding medium caused by the flow of fluid (most applications use air) between the object and the device. The main advantage of using Bernoulli end effectors is that they can perform contactless manipulation of parts. However, they are limited to mostly flat parts, so they are often used to contactlessly handle integrated circuit wafers in different stages of integrated circuit production [2-8], to manipulate large flat lithographic plates in the printing industry [9], and to manipulate hot glass...
sheets in the production of vehicles [10]. Here, after redesigning Bernoulli end effectors to maximize their lifting force, a novel mode of operation is introduced that greatly expands the capabilities of these devices.

2. Previous Work

This study of Bernoulli end effectors was inspired by a practical problem in an air jet based active tooling for vibratory bowl feeders [11]. When an air jet was used to reorient small parts in a vibratory bowl feeder, regardless of the air jet pressure, a number of parts failed to reorient when the pressurized air was applied against them. The airflow between the bowl wall and a part created an attractive force between them large enough to actually press the part against the bowl wall. The slight curvature of the bowl wall increased this attractive force. This problem caused by the Bernoulli effect was solved empirically by separating the parts from the bowl wall by mechanical means. The separation distance was also determined experimentally.

A classification of contactless part manipulation using different levitation techniques is presented in [12]. In fluid mechanics literature [13, 14], Bernoulli end effectors are described as examples of axi-radial flow or diverging radial flow. Paivanas and Hasan [14] assume one-dimensional, steady, laminar, incompressible flow conditions for diverging radial flow between two parallel closely spaced disks. Their theoretical results on spatial pressure distribution and resultant disk force closely match experimental findings except in the region where the axial flow changes into radial. Erzincanli and Sharp [15] propose Bernoulli end effectors for non-contact manipulation of non-rigid materials like food products, while design and operational considerations of Bernoulli end effectors are presented in [16]. A simplified force analysis assuming flow conditions as in [14] using a modified Bernoulli end effector comprising a deflector and a set of axially placed ribs is presented in [17] where the end effector is used for contactless manipulation of sliced tomatoes and cucumbers. Pressure distribution, air velocity, and the gap size are calculated using momentum conservation and common friction factors under assumptions of turbulent and compressible flow [18]. Vortex levitation devices using swirling air flow are proposed as a variation on Bernoulli end effectors in [19].

3. Principles of Operation

Figure 1 depicts perspective and front views showing a typical Bernoulli end effector handling an object having at least one flat surface facing but not touching the plate of the end effector. Figure 2 presents perspective and front views showing a Bernoulli end effector handling an irregularly shaped object. The end effector consists of a gas-carrying tube with a nozzle attached to the center of a plate. The bottom surface of the plate is flat.

To manipulate an object, a Bernoulli end effector is placed close to the object or in contact with it. Then, a gas flow is established through the tube and out of the nozzle of the Bernoulli end effector. The pressure of the gas directly hitting the object creates a force pushing the object away from the end effector. This force is prevalent when the object is at least a few millimeters away from the end effector. However, when the object is closer to the end effector, the radially moving air creates a negative pressure differential between the plate of the Bernoulli end effector and the object creating an attractive force between the two (the Bernoulli effect). For the correct operation of the device, this force (actually the difference between the Bernoulli effect force and the pushing force) must be high enough to move the object when the end effector is moved. The gap between the plate and the object remains during object manipulation allowing non-contact manipulation of the object by the Bernoulli end effector. At the end of the manipulation, the object is released from the end effector by switching off the gas supply to the tube.

During non-contact manipulation of an object, the manipulated object has at least one flat surface facing but not touching the plate of the end effector. During the manipulation of an object when the manipulated object is an irregularly shaped object, the object is in contact with the end effector. Either concave or convex object surfaces can be in contact with the end effector. As an additional advantage, the presented process can manipulate objects having porous or solid surfaces in both the contact and non-contact modes.
An Experimental Analysis of Bernoulli End Effectors for Contact Manipulation of Irregularly Shaped Objects

By adjusting one or more process conditions such as the mass flow of gas, the bottom plate surface type, shape and area, the nozzle shape and dimensions, etc., the process can be optimized for maximum attraction force between the plate and the object. The desired force distribution on the object can also be designed.

The air flow between the lower plate surface and the object surface facing the plate is turbulent and compressible in a region radially close to the nozzle. According to [17] and [18], the resultant force between the plate and the
object depends on the pressure at the nozzle, the cross sectional area of the nozzle, the size of the gap, the radial dimension of the plate, the square of the radial velocity of the air, the air density, the friction factor [20], and the air viscosity.

The forces occurring during the manipulation process are analyzed qualitatively. As the air flow is established between the object and the end effector plate an attractive resultant force starts acting on the object. At the moment the object is pulled vertically towards the plate of the Bernoulli end effector, the attraction force between the end effector and the object is larger than the weight of the object. This accelerates the object towards the end effector plate. If the object is sufficiently far from the plate the attractive force increases since the gap decreases thus increasing the air velocity. However, as the gap decreases further, the effects of friction increase slowing the air, thus decreasing the resultant attractive force. At some gap size the average attractive force becomes equal to the object weight resulting in object levitation.

In the case of contact manipulation, an object is pulled towards the end effector plate with a force larger than its weight. The gap size is determined by the geometry of the object. Additionally, the created contact allows objects to be tilted since there is friction between the end effector and the object. In the next section, a few process variables will be analyzed experimentally. Also, the versatility of this end effector will be demonstrated by a few representative examples.

4. EXPERIMENTAL RESULTS

In this section, two experimental setups are described. Experiments in contactless and contact mode are performed to study the effects of plate surfaces (surface shape, roughness, etc.) on the maximum attractive force between the plate and the manipulated objects. A few applications illustrating the advantages of Bernoulli end effectors over other end effectors are highlighted.

4.1. EXPERIMENTAL SETUP

There are two experimental setups constructed for this study. The first setup uses a Bernoulli end effector mounted on a stand via a fixture. The height of the Bernoulli end effector can be adjusted by sliding the fixture up and down the stand. Once the desired height is reached the Bernoulli end effector can be fixed in the desired position by a screw on the back of the fixture. The end effector is further connected to shop air via a filter lubricator regulator (FRL) unit, a pneumatic on/off valve, and a pneumatic tube. The regulator is adjusted to a desired air pressure. Two aluminum, 150mm x 150mm x 15mm, rectangular plates are used, one with a flat bottom surface and the other with a concave bottom surface. In the second setup, two 140 mm-diameter cylindrical plates (one with a flat bottom surface and the other with a concave bottom surface as shown in Figure 3) are used without the stand to test friction effects in contact mode.

4.2. EXPERIMENTAL SET 1: CONTACTLESS MANIPULATION STUDIES

The following examples describe non-contact manipulation of objects that have at least one flat surface. The experimental setup with the stand is used. According to Figure 1, the end effector is placed at the top of and close to the object to be manipulated. The gap size between the plate surface and the flat surface of the manipulated objects is
An Experimental Analysis of Bernoulli End Effectors for Contact Manipulation of Irregularly Shaped Objects

in the order of 1 mm. Then, by switching the pneumatic valve on, pressurized air is applied through a 6.35-mm inner diameter gas-carrying tube of the end effector. By varying the air pressure from 280 kPa to 550 kPa, cylindrical objects of various sizes and weights (from 180 to 1200 grams) are lifted.

Then, the Bernoulli end effector is modified to maximize the force between the bottom surface of the plate and the manipulated objects. The bottom surface of the plate is modified in such a way to create a concave surface of various concavity angles. The angles of concavity from 0º to 9º in increments of one degree are investigated. To create these concave surfaces, the aluminum plate of previous runs is secured in a lathe and a shallow cone at an angle of about 1º with respect to the original flat bottom surface is cut out from the flat bottom surface. The cone vertex is in the center of the plate and the center of the nozzle. Then, this plate is used to lift objects of various weights at various pressures. The process with a modified plate up to about a 9º-cone cutoff is investigated. The maximum force is obtained for about a 3º angle of concavity. In this case, by varying the air pressure from 280 kPa to 550 kPa the end effector is able to lift cylindrical objects from 230 grams to 1400 grams - resulting in a 20% increase in lifting capacity.

The velocity of the air changes from axial (straight down) to radial (horizontal) as the air hits the object. Assuming the velocity vector only changes the direction, and assuming conservation of mass flow, the area of the nozzle and the cylindrical area just outside the nozzle and between the end effector and the object must be equal, i.e.

\[
\pi r_0^2 = 2\pi r_0 h_0
\]

where \( r_0 \) is the radius of the nozzle and \( h_0 \) is the starting gap size. From Equation (1) the starting gap size is

\[ h_0 = r_0 / 2. \]

Calculating starting gap sizes from the concavity angles of 2º, 3º, and 4º for a 140 mm-diameter Bernoulli end effector leads to 2.45 mm, 3.67 mm, and 4.88 mm starting gap sizes, respectively. However, calculating the starting gap size of the experimental setups using Equation 2 is not straightforward since the nozzle is only partially threaded into the end effector plate. This means that the nozzle is not in the plane of the plate. It is recessed about 5 mm. The outer radius of the nozzle is 9 mm. So, for \( r_0 = 9 \text{ mm} \), \( h_0 = 4.5 \text{ mm} \). This result matches (within the resolution of the experiment) the experimentally obtained starting gap size for maximum resultant force.

Another set of experiments is performed to investigate the effect of the surface finish of the bottom surface of the end effector plate on the force between this surface and the manipulated objects. End effector operations with the aluminum plate used previously having a rough surface, a machined surface, and a manually polished surface are compared. The manually polished surface exhibits the highest force confirming that friction cannot be neglected.

To prove the capability of the described Bernoulli end effector to manipulate porous objects, a 216 mm x 115 mm piece of Vector Electronic Company punchboard with 1 mm diameter through holes arranged in a grid with a hole-density of 16 holes per cm² is successfully lifted.

4.3. EXPERIMENTAL SET 2: CONTACT MANIPULATION OF IRREGULARLY SHAPED OBJECTS

This example describes the manipulation of irregularly shaped objects using a Bernoulli end effector in the contact mode. As in previous examples, the end effector of Figure 2 is connected to shop air via an FRL unit, a pneumatic on/off valve, and a pneumatic tube. The regulator is adjusted to a desired air pressure. The end effector plate, with dimensions of 150mm x 150mm x 15mm, made of aluminum and having a flat bottom surface is mounted on the stand. According to Figure 2, the end effector is placed on the top of the object to be manipulated. Then, by switching the pneumatic valve on, pressurized air is applied through a 6.35 mm inner diameter gas-carrying tube. In this example, a number of irregularly shaped objects of various shapes, weights, and sizes are manipulated. The force created between the bottom surface of the plate and the object due to the Bernoulli effect allows manipulation of the object by moving the end effector. In addition, irregularly shaped objects from about 100 grams to about 200 grams can be tilted from 4º to 11º using air pressures from 280 kPa to 550 kPa.

When a concave surface having a concavity angle of about 3º is used as the bottom surface of the end effector plate, the force between the bottom surface of the plate and the object is increased. During manipulation, irregularly shaped objects from 100 to 200 grams can be tilted from 10º to 35º using air pressures from 280 kPa to 550 kPa, which is about a 24º increase in performance at high pressures, or 300%. 

Page 448
4.4. APPLICATIONS

Contrary to the current opinions on the limited uses of Bernoulli end effectors, Figure 4 illustrates the wide range of objects that can be manipulated with these devices. In Figure 4, a 45.7 gram plastic bottle is manipulated in the contact mode. The bottle is picked up from all possible sides, on the side as shown in Figure 4a, from the top as shown in Figure 4b, and from the bottom as shown in Figure 4c. For “grabbing” the bottle on the side and the bottom a pressure of about 140 kPa is used. Since there is a rather limited surface area on the top of the bottle, for “grabbing” the bottle from the top a somewhat higher pressure of 180 kPa is applied. It is interesting, in this example, that all contact surfaces are convex. Next, figure 4d depicts a view of a rock (an irregularly shaped object) manipulated by the Bernoulli end effector. The rough surface of the rock would present a problem for vacuum chucks, and its irregular shape would pose a problem for many mechanical grippers. Picking the bottle from the top would also be difficult for many classical end effectors.

![Figure 4: Bernoulli end effector applications: (a) holding a bottle in a horizontal position; (b) holding the same bottle in an upright position; (c) holding the bottle upside down; (d) holding a rock](image)

Advantages of the contact mode of manipulation over the contactless mode are illustrated in Figure 5. The plastic bottle is tilted about 60° as pictured in Figure 5a and 90° as depicted in figure 5b. In contactless mode, since the manipulated object is resting on an air cushion there is little friction between the end effector and the object. This
means that objects could slip easily from the end effector. In contact mode, the Bernoulli attractive force acts like glue holding the object against the plate of the end effector.

Figure 5: Bernoulli end effector application: (a) tilting a bottle approximately 60º, (b) tilting a bottle 90º

5. SUMMARY

Bernoulli end effectors are analyzed experimentally. A novel contact mode of operation is described and justified. The maximum lifting force of these end effectors, in both contact and contactless modes, is increased about 20% experimentally by turning the flat plate into a cone with a 3º-concavity angle. The range of objects that can be manipulated with Bernoulli end effectors is greatly expanded. This work illustrates and justifies additional advantages of Bernoulli end effectors over their mechanical counterparts in that they can manipulate a wide variety of objects without the need for complicated apparatus or retooling; they can manipulate objects having at least one relatively flat surface without contact with the end effector; they can manipulate objects having porous or rough surfaces; and, in contact mode, they can manipulate and tilt irregularly shaped and/or porous objects having concave, convex, and/or flat surfaces.

REFERENCES

26, 2000.


RFID Systems in Manufacturing SMEs – Benefits and Limitations

Sara Zarei*, Simon Hodgson¹, David Ford² and Antony Robson²

¹ School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA, UK

² Stanley Vickers Ltd
Snowdon Road
Middlesbrough, TS2 1LG, UK

ABSTRACT

Radio Frequency Identification (RFID) is an AIDC technology that is used to identify people or objects carrying encoded microchips through the use of radio waves. RFID technologies were initially developed by the British military in the 1940s in order to be able to distinguish their own aircraft from the enemy aircraft by equipping the aircraft with devices to send out coded signals. RFID technologies have since found applications in many everyday devices such as remote controls for gates or garage doors, anti theft devices in vehicles, security systems and the new European Union passport as well as becoming widely used within the manufacturing industry to allow information to pass easily and accurately between different stages of the supply chain. RFID technology has been successfully implemented within many large multinational organisations but research into the successful implementation of RFID systems within SMEs is far more limited.

An investigation was conducted at a Teesside manufacturing SME to identify the areas in which the implementation of an RFID system would benefit the company through improved efficiency and visibility of the processes and also the accuracy and reliability of real-time data transfer. An exploration of various RFID systems and their function within the manufacturing environment was undertaken so as an appropriate technology could be identified that may be suitable for implementation within the SME to improve the manufacturing process. The investigation of a feasible AIDC technology is then outlined and the determination of whether this technology may add value to the current operations at the SME is analysed. The possible limitations of the success of an RFID system and also any problems that may occur either during implementation stage of for future performance were identified alongside the perceived benefits for the company including a cost benefit analysis.

1.0 INTRODUCTION

In the manufacturing industry, the tracking of a product and its status through the processes of a company, from the moment it arrives until the time of despatch, can significantly reduce the lead time, increase the efficiency and visibility, increase the accuracy of production plans and also improve the efficiency of an Enterprise Resource Planning (ERP) system. As the constant demand for total visibility and fast and accurate information flow increases, the inefficiencies of traditional systems are becoming more apparent. [1]

Automatic Identification and Data Capture (AIDC) technologies and systems were initially developed as a response for the requirement to collect and transfer levels of data in an efficient, quick and accurate manner. AIDC technologies can reduce or even eliminate the occurrence of errors in the collection of data and the manual input or entry of data and has been successfully applied in the retail, logistics and manufacturing industries. Some AIDC technologies, such as barcodes, have found widespread applications in the manufacturing industry within material and tool control, shipping and transportation. [1 – 3] However, the potential uses of RFID within the supply chain have only recently been recognised even though the technology has been available for over 60 years. The increased data capacity and flexible characteristics of Radio Frequency Identification (RFID) technologies have highlighted them as a potential answer for solving these problems and in turn have caused interest from world-class
organisations in their aim to increase visibility [4 – 5]. RFID technologies are expected to offer major savings for an organisation by providing the ability to identify individual items with ease and track them accurately throughout the supply chain, distinguishing their location and movements at all times. However, the substantial investment required for the implementation of an RFID system and the practical challenges and cultural resistance to change that may follow have caused controversy in the true value of the technology. [1][2][6].

It is often thought of as complicated, expensive and time consuming to adopt a new technology project within a company and this is often the reason why many are reluctant to alter their methods. RFID is a technology that is very much tarnished by this assumption and so the real benefits are often overlooked. The scale of implementation and the specific requirements of the system will dictate the investment value required and can vary significantly between cases, so a set cost and benefit assumption for all is impossible. [7] However, the centre for Auto-ID has now established a calculation method for the Return on Investment (ROI) of an implemented RFID system in order to educate companies on the cost benefits associated. [8] Although this may significantly help many companies, the limitations of the calculator often mean that an inaccurate quote is provided. The cost of a tag and the cost of the item to be tagged have limitations set which means that it is often unable to cater for all organisations.

This paper aims to educate and assist organisations with the benefits and limitations of RFID systems within the manufacturing industry, and in particular within SMEs. A method of assessing the cost benefits of an RFID system application has been previously researched by the authors and this case study will be used to support the conclusions made. There is limited research into the adoption of RFID systems within manufacturing SMEs, especially with regards to cost benefits, and so the outcomes of this study aim to assess the true value of this technology within the SME manufacturing field.

2.0 Methodology

The methodology for quantifying a cost benefit analysis outlined in this paper is based on the notion of waste minimisation and lean manufacturing complementing and involving the application of a suitable AIDC technology. Initially, the current manufacturing processes at the SME and the flow of materials and information across the shop floor for each part produced were analysed through a process mapping investigation previously conducted by the author [9]. The aim of the process mapping exercise was to eliminate any operations in the process that are classed as non-value adding. A non-value adding operation is any activity that does not directly add value to the end part or process, or an activity that is classified as a waste. Non-value adding operations should always be attempted to be eliminated from the process, however, sometimes they may be an operation which must be performed for the process to flow but does not directly add value. Such operations cannot be fully eliminated from the process as they are required for the successful completion of the product or service. [10 – 11] The future state manufacturing processes could then be mapped out, taking into account the improvements identified through lean thinking and the implementation of AIDC. Lastly, a cost benefit analysis is conducted in order to weigh the benefits of implementation against the associated costs of implementation.

Once the process mapping exercise was conducted for the current and visualised state of the manufacturing process for an example of each product, a cost saving analysis can be carried out to assess the potential savings that can be made. In order to analyse the savings as accurately as possible, some information and figures were obtained from the company as shown in Tables 1 and 2. The annual sales figures, inventory and tracking data were collected in order to evaluate the cost benefits through improvements.

Table 1 – Annual Sales

<table>
<thead>
<tr>
<th>Annual New Screw Sales</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual New Barrel Sales</td>
<td>150</td>
</tr>
<tr>
<td>Annual New Feed Liner Sales</td>
<td>80</td>
</tr>
<tr>
<td>Annual General Engineering Sales</td>
<td>50</td>
</tr>
<tr>
<td>Annual Screw Refurbishment Sales</td>
<td>350</td>
</tr>
<tr>
<td>Annual Barrel Refurbishment Sales</td>
<td>50</td>
</tr>
</tbody>
</table>
RFID Systems in Manufacturing SMEs – Benefits and Limitations

**Table 2 – Inventory and Tracking Data**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventory Value (£)</td>
<td>130,000</td>
</tr>
<tr>
<td>Safety Stock (%)</td>
<td>50</td>
</tr>
<tr>
<td>Carrying Cost (%)</td>
<td>10</td>
</tr>
<tr>
<td>Average Labour Cost Per Hour (£)</td>
<td>10</td>
</tr>
<tr>
<td>Number of Times Physical Tracking is Performed</td>
<td>52</td>
</tr>
<tr>
<td>Time Taken to Complete Physical Tracking</td>
<td>5</td>
</tr>
</tbody>
</table>

The inventory levels kept at the company can be classed as a waste because a significant amount of money is tied up in safety stock. The level of safety stock could be significantly reduced through the implementation of an AIDC system as it can provide 99.92% accuracy through the supply chain [12 – 13]. The potential savings involved for the increased accuracy and in the reduction of inventory through the implementation of an appropriate AIDC system may be calculated using the following equation:

\[
\text{(Total Inventory Value)} \times \text{(Carrying Cost (%))} \times \text{(Safety Cost (%))}
\]

Manual counting times are also classed as a waste as they take a significant amount of time to complete. Manual counting times may be reduced by 80% through the use of an AIDC system [12 – 13] and the savings involved in this reduction of time may be calculated using the following equation:

\[
\text{(Number of times a manual count is carried out)} \times \text{(Time taken to complete a manual count)} \times \text{(Hourly labour cost)} \times 80\%
\]

The process mapping exercise, previously completed by the authors [9] was used to critically analyse the current state of the operations performed in order to establish areas of waste and identify areas that could be improved through the implementation of an AIDC system. The modified state was then visualised based on these improvements for each process and the financial benefits were outlined.

The process efficiency can be drastically increased through the implementation of an AIDC system by the elimination of non-value added operations within the process. In order to calculate the improvements felt in the process efficiency through the implementation of an AIDC system, the cost of the current state process was compared with the cost of the modified state process. This was calculated using the following equation:

\[
\text{(Process duration (h))} \times \text{(Number of parts manufactured per process)} \times \text{(Hourly labour cost)}
\]

Once the process mapping exercise and process improvements were established, an investigation into an appropriate AIDC technology capable of providing these improvements for each process was examined. The suitability of each AIDC technology was examined in order to find the most appropriate system for use at the company. Each technology was considered regarding the areas of equipment size and the feasibility of this equipment with the process at the company, the compatibility of the technology with the products at the company and their surface material, the ability of the technology to function under the operating condition felt by the products manufactured at SVL and also the software requirements for each technology. As all products manufactured at the company are made from metal, an appropriate technology must have the ability to be applied to and perform well on metallic surfaces. The harsh operating conditions such as extreme temperature levels, molten plastics or rubbers, chemicals for cleaning and other general machining that are exhibited in this field of work. A major restraint with for any AIDC system utilised within such a company is the temperature extremities. The technology should also have the ability to be applied to the product without compromising or affecting its performance in any way. These above criteria formed the basis of the investigation into a suitable technology for use within the SME alongside the flexibility of the system itself.
3.0 Feasibility and Application

Through the process mapping investigation previously conducted at the company, it can be seen that the implementation of an AIDC system could significantly reduce the lead times of processes, increase visibility and communication across the shop floor and improve efficiency of the processes. The author has previously identified RFID as the most suitable AIDC technology to be implemented at the company through a feasibility analysis [9] as summarised in Table 3.

<table>
<thead>
<tr>
<th>AIDC Technology</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Barcodes</td>
<td>+ Low cost to implement</td>
</tr>
<tr>
<td></td>
<td>+ Well established</td>
</tr>
<tr>
<td></td>
<td>− Large label of 1.5cm x 9.5cm minimum for level of data required</td>
</tr>
<tr>
<td></td>
<td>− Easily damaged with dirt/grease/scratches</td>
</tr>
<tr>
<td></td>
<td>− Unable to modify data once applied (fixed data)</td>
</tr>
<tr>
<td>2D Barcodes – Direct Part</td>
<td>+ Small label of 1cm x 1cm for data</td>
</tr>
<tr>
<td>Marking (DPM)</td>
<td>+ ‘engraved’ into metal surface of parts</td>
</tr>
<tr>
<td></td>
<td>+ Error correction algorithm to allow data to be read after suffering partial damage</td>
</tr>
<tr>
<td></td>
<td>− Can be damaged through manufacturing and performance of parts</td>
</tr>
<tr>
<td></td>
<td>− Fixed scanning stations to read label limits flexibility/portability</td>
</tr>
<tr>
<td>Magnetic Stripes</td>
<td>+ Data can be modified after application</td>
</tr>
<tr>
<td></td>
<td>− Cannot be applied to metal surfaces as data may be lost or interference can occur</td>
</tr>
<tr>
<td>RFID Tags</td>
<td>+ Can be ‘embedded’ into part for minimal interference in performance</td>
</tr>
<tr>
<td></td>
<td>+ ‘line of sight’ not required to scan then tag</td>
</tr>
<tr>
<td></td>
<td>+ Large amounts of data may be stored and modified as required</td>
</tr>
<tr>
<td></td>
<td>+ Portable hand held readers allow for flexibility in manufacturing environment</td>
</tr>
<tr>
<td></td>
<td>− High implementation costs</td>
</tr>
<tr>
<td></td>
<td>− Metal environments may cause poor signal/communication problems in cases</td>
</tr>
</tbody>
</table>

The selection of a suitable AIDC technology also depended on the requirements of the system itself as it is required to be wireless for use on the shop floor. It also must be decided whether or not the technology will be used for the internal tracking of the part only through the manufacturing process or whether the technology must be present throughout the lifespan of the product for future refurbishment. The size and shape of the products manufactured at the SME must be compared with the size of the RFID equipment in order to establish if they are compatible with one another. A comparison of the available RFID tags available was conducted in order to distinguish those of an appropriate size for the products manufactured. Active RFID tags were considered unsuitable due to their size and the presence of an internal power source as they would affect the performance of the product to which they were applied. Passive RFID tags were further investigated as they are much smaller and do not require an internal power supply. Passive RFID tags are available in various shapes and sizes [14] as shown in Figure 1.

![Figure 1 – Types of Low Frequency RFID Tags [14]](image)

The application of an RFID tag onto the parts cannot be external due to the potential risk of damage to the tag itself and also the interference it may cause with the products functionality. The tag must therefore be embedded into the body of the product to protect from damage and allow the product to function properly. The glass type RFID tags...
RFID Systems in Manufacturing SMEs – Benefits and Limitations

were found to be suitable to apply to the products manufactured at the company as they could be embedded into the parts and were small enough to be discreet and not interfere with their manufacturing or performance. The glass tags come in a variety of sizes, ranging from 12mm to 32mm in length which makes them versatile for parts of all sizes with minimum machining for the application. [15]

A major restraint with regards to the implementation of an RFID system at this company is the temperature extremities, as the screw flights are heated to temperatures of a maximum of 200°C during the plasma hardfacing process, which is too high for the glass tags to handle and would cause them to be irreparably damaged. However, only the screw flights are subjected to this process and so the tag may be placed in a different area of the part to avoid damage through the hardfacing process. If the RFID system is required to be present on the product throughout its lifecycle, in the event of refurbishment being required, then it would be subjected to operating temperatures ranging between 110°C and 320°C. However, there are areas of each product that are not subjected to these extreme temperatures which would only experience temperatures up to a maximum of 90°C and the tags would therefore have to be applied to the area of the product that is exposed to the least heat. For example, Figure 2 shows some screws and barrels manufactured at the company, the drive section of the screw and the section of the barrel before the hopper slot would be the places when temperatures were lowest at all times and so the tags may be embedded here to avoid damage. The RFID system would also be expected to withstand certain levels of oil, grease and dirt along with levels of damage that could occur throughout the manufacturing operations that the product would be exposed to. The application of an embedded glass tag with an adhesive resin for protection has the ability to withstand such conditions. The fact that the tag is embedded within the part would avoid levels of damage through machining and the adhesive resin used with the glass tag would provide a waterproof and protective coating that would not allow grease, dirt or liquids to reach it.

![Figure 2 – Barrels (left) and Screws (right) manufactured at the SME](image)

RFID technologies are known to have some limitations when used with metallic objects as metal interference can significantly reduce the reading range of the tag and also may cause loss of information through interference. Low frequency (LF) glass type RFID tags were found to be the most appropriate solution for this problem as they may be easily embedded and set in place through the use of an adhesive resin, which would provide protection for the tag from grease and oil and also provide a barrier layer between the tag and the metal surface which would allow for the data to be read without any interference.

4.0 **Cost Benefit Analysis**

One of the main barriers to the successful implementation of an RFID system, especially within an SME, is cost. Due to the fact that the current level of research into the successful implementation of RFID within SMEs is limited and the level of education in SMEs with regards to RFID is inadequate; there are many perceptions that the cost of implementation is far greater than the benefits. [7, 16-17] This is also burdened by other factors; such as the implementation of a new technology within a company requires education, support from all workers, sufficient funding and can also be time consuming. RFID technologies are expected to provide major savings for an organisation by providing the ability to identify individual items with ease and track them accurately throughout the supply chain, distinguishing their location and movements at all times. However, the substantial investment required for the implementation of an RFID system and the practical challenges and cultural resistance to change that may follow have caused controversy in the true value of the technology. [16]
There are currently no standard methods known to calculate the ROI for the implementation of an RFID system as it will vary greatly for each company. In order to establish the cost benefits of the implementation of an RFID system within the company, a number of simple calculations can be carried out. The savings that can be observed through the implementation of an RFID system were outlined previously through equations (1), (2) and (3) and the results of these calculations are as follows.

The level of safety stock could be significantly reduced through the implementation of an AIDC system within the company. Using the data collected from the company, Table 2, the estimated savings that can be gained from the reduction of inventory levels through the implementation of such a system are as follows:

\[
\text{Savings} = (\text{Total Inventory Value}) \times (\text{Carrying Cost} \%) \times (\text{Safety Cost} \%)
\]

\[
= (£130,000) \times (10\%) \times (50\%)
\]

\[
= £6,500
\]

The cost of manual counting and tracking times can be reduced by 80% through the implementation of an AIDC system. The manual tracking of the jobs across the shop floor was completed on a weekly basis and took 5 hours in total to complete, including the manual input of the results. Using the data seen in table 2, the estimated savings that can be seen through the reduction in the costs of manual tracking can be seen to be:

\[
\text{Savings} = (\text{No. of counts}) \times (\text{Time taken per count}) \times (\text{Hourly labour cost}) \times (80\%)
\]

\[
= (52) \times (5) \times (10) \times (80\%)
\]

\[
= £2,080
\]

The process efficiency may also be increased through the implementation of an AIDC system through the improvements stated in the process mapping exercise outlined. In order to calculate the predicted saving improvements in the process efficiency through the implementation of an AIDC system, the cost of the current state process must be compared with the cost of the modified state process, which is calculated using equation (3). The elimination of the non-value added processes allowed a time saving to be measured for each process which was used alongside the data from tables 1 and 2. The amount of hours saved for each process were then used to calculate the cost savings per process. The annual cost savings could then be calculated by multiplying the number of processes per year as seen in table 4.

**Table 4 – Cost Benefits for Each Process**

<table>
<thead>
<tr>
<th>Process</th>
<th>Hours saved per process</th>
<th>Cost savings per process (£)</th>
<th>Cost savings per year (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Screw</td>
<td>8</td>
<td>(8) x (1) x (10) = 80</td>
<td>(80) x (300) = 24,000</td>
</tr>
<tr>
<td>New Barrel</td>
<td>3</td>
<td>(3) x (1) x (10) = 30</td>
<td>(30) x (150) = 4,500</td>
</tr>
<tr>
<td>New Feed Liner</td>
<td>6.25</td>
<td>(6.25) x (1) x (10) = 62.5</td>
<td>(60) x (80) = 5,000</td>
</tr>
<tr>
<td>Tie Bar</td>
<td>7.75</td>
<td>(7.75) x (1) x (10) = 77.5</td>
<td>(77.5) x (50) = 3,875</td>
</tr>
<tr>
<td>Screw Refurbishment</td>
<td>5.5</td>
<td>(5.5) x (1) x (10) = 55</td>
<td>(55) x (350) = 19,250</td>
</tr>
<tr>
<td>Barrel Refurbishment</td>
<td>4</td>
<td>(4) x (1) x (10) = 40</td>
<td>(40) x (50) = 2,000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td><strong>£58,625</strong></td>
</tr>
</tbody>
</table>

The improvements in the process efficiency and transportation times in particular can be identified through the use of process activity mapping and flow across the shop in order to add value to the operations and also savings in the reduction of safety stock levels and manual part counting can also be identified. The total annual saving that
were calculated for the company across each of these areas through the implementation of an RFID system and the improvements in the shop floor layout are outlined in Table 5.

Table 5 – Annual Savings

<table>
<thead>
<tr>
<th>Savings</th>
<th>Amount (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventory Savings</td>
<td>£6,500</td>
</tr>
<tr>
<td>Manual Tracking Savings</td>
<td>£2,080</td>
</tr>
<tr>
<td>Process Efficiency Savings</td>
<td>£58,625</td>
</tr>
<tr>
<td><strong>Total Annual Savings</strong></td>
<td><strong>£67,205</strong></td>
</tr>
</tbody>
</table>

The process efficiency savings were calculated through the improvements to the operations in each process and the shop floor layout across all products produced annually. In total, the recommended alterations and implementation of an integrated RFID system has the ability to save the company an estimated £67,205 per year.

The cost of the RFID tags themselves has been previously identified as a major disadvantage of the technology [16 – 17] and was therefore investigated with respect to the requirements at the SME. The glass tags that were identified as the most appropriate technology for the application of an AIDC technology within a metal part that is exposed to harsh working conditions were found to range from £1.50 to £2.80 per tag dependant on the size required. The cost of these tags was found to be negligible with respect to the value of the products manufacture at the company. The handheld RFID reader devices were found to range from less than £500 up to over £2000 pounds for a pair including a power cradle [18 – 19]. The appropriate reader system required would need further investigation and again depends upon the level of interference felt from the metal environment, the reading distance required and the facilities required from the device. The installation and integration costs of an RFID system were found to range between £23,000 and £25,000.

The investment necessary for the implementation of an RFID system is highly dependant on the requirements of a company and the hardware chosen. The implementation costs for the SME were initially established as summarised in Table 6.

Table 6 – Implementation Costs

<table>
<thead>
<tr>
<th>System Requirements</th>
<th>Associated costs</th>
<th>Investment Required (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags - £2.00 Each</td>
<td>580 required for ‘new sales’</td>
<td>£1160 (per year)</td>
</tr>
<tr>
<td></td>
<td>(580) x (2) = 1160</td>
<td></td>
</tr>
<tr>
<td>Extra Labour Required for Tagging</td>
<td>30 minutes allowance for each tag application</td>
<td>£2900 (per year)</td>
</tr>
<tr>
<td></td>
<td>(Time taken for each tag (hrs)) x (hourly labour (£)) x (No of tags)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5) x (10) x (580) = 2900</td>
<td></td>
</tr>
<tr>
<td>Hand held Readers</td>
<td>(£2000 per pair) x (6 required)</td>
<td>£12000 (one purchase)</td>
</tr>
<tr>
<td></td>
<td>(2000) x (6) = 12000</td>
<td></td>
</tr>
<tr>
<td>Software and installation costs</td>
<td>£25000 max</td>
<td>£25000 (one purchase)</td>
</tr>
<tr>
<td><strong>Total Initial investment required</strong></td>
<td></td>
<td><strong>£41,060</strong></td>
</tr>
</tbody>
</table>

The initial investment required for the implementation of a suitable RFID system has been calculated as £41,060 with an additional cost of £4,060 per year to tag each new product manufactured. The ROI of the RFID system could therefore be calculated by dividing the cost benefits of the system with the investment required to implement the system.
ROI = (Cost benefits of system (£)) / (System implementation costs (£))

= (67,205) / (41,060)

= 163.68%

The ROI for the implementation of an RFID system within this company was found to be 163.68% which means that the payback period is about 8 months.

5.0 SUMMARY

The major barriers for the successful implementation of an AIDC system within a company are found to be: the lack of basic IT infrastructure, the reluctance to invest in new technologies due to lack of education and knowledge in the subject area and cultural resistance to change. A major issue for the implementation of an AIDC system within an SME is the requirement of a computerised infrastructure as a prerequisite as this is associated with high costs and vast amounts of training. [20 – 22]

The accurate traceability of products throughout the manufacturing process can improve the efficiency of a process, the accuracy of the information available and improve the lead time of the product. The traceability of parts may be improved through the implementation of an RFID system with components that are capable of withstanding the conditions of the part. Passive LF glass type RFID tags were identified as a viable solution as they may be embedded into the metallic parts produced at this company and can withstand all conditions felt through the products lifecycle.

The cost benefit analysis for the SME through the implementation of an RFID system and the modification of the shop floor plan was found to provide a saving of £67,205 annually with a payback period of around 8 months. Significant cost savings are found through the implementation of an RFID system based on the reduction of safety stock, the elimination of manual job tracking and the reduction of the manual input and written data throughout the process. The modification of the shop layout better facilitated the use of an RFID system as well as improved the efficiency of the process by providing increased visibility and a large reduction in transportation times.

ACKNOWLEDGEMENTS

The author would like to thank all of the employees at the SME for their support throughout this case study.

REFERENCES

[Accessed November 2008]


[16] Smith, I. G; 2007. RFID – Blessing or a Curse for SMEs. European Centre for AIDC, United Kingdom. Presented at the Conference and Exhibition for RFID.


A generic Framework for Life-Cycle-Management of heterogenic Building Automation Systems

Konstantin Krahtov, Sven Rogalski, Daniel Wacker, Dan Gutu and Jivka Ovtcharova

Process and Data Management in Engineering / Intelligent Systems and Production Engineering
FZI Forschungszentrum Informatik
Karlsruhe, D-76131, Germany

ABSTRACT

The ongoing development of technologies has in recent years lead to a significant increase in the available technical equipment of private as well as commercial buildings, which makes a computerized monitoring and control of mechanical, electrical and energy systems, possible. Known as building automation products and systems that lead to a significant improvement of individual life, safety and working conditions are all combined in this term. With the introduction of building automation techniques like function cycles in systems automatically enforced according to preset parameters make the servicing and monitoring easier. However, products from different manufacturers have to work in any combination with each other, this is very beneficial for customers, but from the perspective of manufacturers and system integrators it poses major problems. The reason for this is the heterogeneity of the products and the difficulty to implement interoperability. During the life cycle of a building it comes repeatedly to changes in the buildings technology for example. This begins with the planning phase and goes on to the commissioning and the continued use of buildings, the space requirements change, leading to a new configuration of the initial automation and manual intervention within the different levels of communication is necessary. Such changes are particularly burdensome especially when a large number of different components from different systems are affected, resulting in a considerable requirement of resources due to the multiple entry of data, inconsistency and incomplete documentation because of knowledge islands. Thus even small changes evolve to complex reconfigurations. The following article deals with a software framework that outlines this problem and allows major simplifications in the administration and modification of complex spatial configurations in the building automation. It shows how digital 2D building plans can be evaluated so that geometric data contained on automation components are automatically read-out so that they can be managed with total control from the central framework.

1. INTRODUCTION

With the construction of a new building, e.g. a family home, various sections of the building automation systems are installed and in operation. Despite standardizations on the communication level, by BACnet or LON, this is currently done with a relatively high cost, as these heterogeneous systems / devices must communicate with each other. On the side of the system integrators (craft industry) it takes know-how and considerable time to implement the necessary control functions. Subsequent extensions of automation components, such as the integration of new sensors or actuators, nowadays often require special solutions, in which problem-specific extensions are programmed, most of the time other applications besides the original are not usable. The user has to use decentralized control elements to those components with isolated features, for this often special skills are required, such as in the management of the heating system. The consequences of this are that these media and the process breaks, spread over the entire life cycle of automation systems in buildings starting from planning, through its commissioning, management and maintenance, through to its disposal (cp. figure 1).

Another challenge in the field of energy-efficient building is total integration of several automation systems. To do this they need to adapt to their environment by using the "environmental job," for example, air conditioning, lighting or energy production, use and freely available resources. System integrators will need an in-depth knowledge, ranging from components and wiring in the physical level to domain-specific functions and the abstract connection specific functions of various trades on the logical level. For this reason, new methods are needed; a simplified, manageable and cross-system building and configuration that allows easy small projects management by small and medium-sized systems integrators.
The here outlined problems have been met in the past with various approaches to a largely automated support of the life cycle of building. There has already been put some effort in the standardization of the data protocol to ensure that devices from different manufacturers can be combined, examples for this are BACnet and LON. However, these efforts stay behind their expectations because of various problems (see [1] [2]).

In the interest of geometric and functional planning of the building (GA) in the engineering phase up to this date various software tools available. For the geometric design of a plant 2D or 3D CAD programs can be used. Most are AutoCAD based and cover all domains of the building, such as ventilation, heating, lighting, access control, photovoltaics, such as "Rocade" of Mensch und Maschine AG [3]. In addition, there is specific software such as "Elaplan" of ElektraSoft [4], "Vaillant WinSoft" of the company Vaillant [5] or "Simatic Automation Designer" by Siemens [6] etc. offers additional features for building systems, such as inventory management, cable laying and generating schedules of images for monitoring and control. They work on CAD-based or have interfaces to CAD systems and standard calculation programs with import/export functions. However, these systems are confined to the data generation and management, and generally have no intelligent combination of data types. Therefore they do not allow for seamless transfer of geographic positioning of the components and cover only the planning stage.

The monitoring and control of a facility is realized with so-called S.C.A.D.A. * tools, covering a wide range of programs for monitoring and control of a GA facility. Some S.C.A.D.A. systems already provide interfaces to CAD systems. "WizCAD" for example converts with the module "Wizcon Supervisor" [7] AutoCAD plans to Wizcon S.C.A.D.A. -specific files. In "PVSS II" [8] CAD plans are converted to XML-based system PVSS pictures. These are application-specific interfaces, which are usually based on proprietary mechanisms. Additional extensions or linkages with other systems, are if the exist at all, limited.

In the area of maintenance and alteration within the Building Automation, special software tools from the realm of facility-/building management is used. Such software evaluates the disruptions caused by coupling with management/building automation, oversees operations for routine preventive measures and has interfaces to software management products. Thus, e.g. the staff at alarm cases can assess the priority for messages and alert the fire service or the service disruption. The system can also trigger the ordering process, when not enough parts are available. A major disadvantage is that there are elaborate procedures for the configuration and implementation of such mechanisms, because of the data exchange of the system with other applications via proprietary interfaces.

Overall, it is observed that isolated solutions for open, cross-system communication and functionality of the systems with automated setup, configuration, modification and coordination of building systems, are currently available only in the research environment. For example, in the project "VIMP" at the University of the Bundeswehr

* S.C.A.D.A.: Supervisory Control and Data Acquisition
Munich with "IMOS home" a tool for planning, commissioning, visualization and operation of an intelligent houses is developed. The linkage of planning with the S.C.A.D.A. functionality is "IMOS" in an interactive graphical user interface with 2D virtual realistic 3D objects is realized [9] [10]. However, this is a concept for systems with a manageable number of components. For the private sphere of life and industrial applications with a large number of nodes, it is not suitable. Furthermore, it is questionable to what extent additional specialist knowledge on the part of the system integrators due to the 3D visualization is necessary. The same applies to that of program InnoNet sponsored project AUTEG, which is trying to design an automated process of building a complex investment [11].

2. REQUIREMENTS OF THE FRAMEWORK

The framework has the ambition of a generic approach to the lifecycle management of automation systems. Therefore, its development has been made in the following two main applications, as well as numerous scenarios of building automation and energy management, which later led to the validation of the results:

• **Model Transformation to configure third-party systems:** The framework is an interpretation of 2D drawings and the generation of object models in formats allows any target, so that the drawings implicit information decrypted by third-party systems are used. This will allow these 2D planning documents as a basis for other configuration systems that are reducing the time and resources for the manual input of data costs. Furthermore, this approach allows for configurable model transformation consistent, redundancy-free and full harmonization of data from different life stages of automation systems with the condition that any changes only in the 2D models are implemented.

• **Generic monitoring and control of heterogeneous automation systems:** An Internet user is the one all interpreted data based on a generic model and visualize rudimentary functions to interact with users. But it should be possible to use data via pre-configured WebServices on this surface. Through this way it will be possible to monitor and control actions to heterogeneous types of automation systems on generic basis and thus an integration of these systems at the level of user interaction.

Based on these cases and scenarios, the functional, technical and management requirements are derived. The main concern of a configuration based adaptation of the framework for different application domains without the need to implement complex functions as well as on the intuitive and user-friendly, respectively efficient interaction with the end user.

2. CONCEPT OF THE FRAMEWORK

Based on the above-described problem for small and medium-sized enterprises (SMÉ), a suitable software framework of FZI Forschungszentrum Informatic supports the lifecycle of buildings fully. It simplifies the management and modification of complex spatial configurations, by using digital 2D AutoCAD building plans (.DXF or .dwg) geometric data contained on automation components in terms of their control are automatically interpreted. The information obtained represents the spatial configuration of the overall automation in a building and can be on the framework-specific function possibilities using database, XML or Web service interfaces. In this way, it is possible for any PC-based or embedded control units to expand, resulting in an automatic transfer of these configurations. In addition, the framework ensures a uniform spatial configuration of automation systems for visualization and rudimentary communication with the control units via Web Services. An open technologies (Java, Eclipse, GEF etc.) based viewer interprets this data on demand in formation and services easy to control, for example, all energy-related components from a 2D drawing planning to extract and visualize. For the users of the framework is thus the ability to communicate information about such components (such as current performance of the photovoltaic component, total electricity consumption, the state of window sensor) is retrieved.

2.1. COMPONENTS OF THE FRAMEWORK

The framework consists of five core components that are either used together or individually in different applications in the field of information management automation and energy systems. Below the individual modules are briefly described (cp. figure 2):

The AutoCAD Interface provides an interface to export 2D AutoCAD drawings into a specially adapted to the requirements of the Framework XML data format. The exported, geometric components can be configured so that
only the necessary information for the subsequent rule interpretation is available. It can also include data from other systems within the framework over the XML interface if the adapters are available.

The Schema Editor as an Eclipse RCP application allows the creation of information objects for the monitoring and control of building automation, without an implementation. For each object properties of different types of data and references to Web services, rules for the interpretation of 2D drawings and the SVG graphical metaphers can be added.

For rule interpretation or semantic enrichment of the CAD system exported data, the Rule Engine is used. The complexity of the generic approach for the adoption of spatial configurations from differently structured CAD drawing data as well as the requirement for transparent and user-friendly configuration make the use of exciting rule technologies impossible. The adjustment flexibility and the clarity of the interpretation of logic can only be customized with a scripting language. The interpretation process is thus dependent on the structure of each the evaluated data from each drawing. In addition, this script is characterized by a dynamic language typing, thus an additional definition of a schema with data type formats to interpret the underlying XML data is nor required.

The backend database connectivity and Trans Pain ter Framework manages all interpreted data, creates an abstraction layer between the database and object-oriented programming language used and adjusts the communication with the frontend. To minimize the effort on the client side, "Painter" classes are provided by the server. A Painter class provides a Java class that contains the instructions for drawing an object. It implements a special interface, which the client (Web browser) draws to instances of this class directly to the object to a Java2D Graphics instance. By keeping the technology within the framework the web browser barely needs libraries to display and a very low volume of data transfer via the Internet occurs.

A central element of the Framework is the front-end component Front end 2D represent ation Java2D appl et, which allows the visual presentation. The presentation is provided through Remote Method Invocation (RMI) by the backend. In order for the Java2D appl et to load as quickly as possible, it is the Trans-Framework Painter from the backend. In addition, large amounts of data are visualized as powerful information through dynamic loading.

For the presentation of the logical topology of the automation and energy systems and for the attributes and WebService various components the Frontend GWT component is used.

Using a Web services interface allows data from third-party systems to the generic web interface from the Framework and control instructions to third-party systems using Web services. From a WSDL file selected configuration of the Web is with the object model of the atomisation / energy system linked so that the correct Web services on the surface of the context are represented.

Figure 2: Framework Architecture
2.2. DATA MODEL OF THE FRAMEWORKS

The heart of the framework described here is the generic data model, which is a configuration instead of a deployment of the types of information objects. The data model includes three main groups (see figure 3):

- the domain model for information about the environment in the framework of the ECS
- the logical model as an instance of the domain model with concrete objects (nodes) and their attributes
- graphic model for the visual representation of the objects

The domain model serves as a meta model to types of real or abstract objects. Here the types of attributes, Web services references and interpret rules are defined. Optional is a relationship with a view (form) in the graphic model, to be able to define how objects of each type are presented. For example, a type "fan" with an assigned representation, as well as an abstract type "channel" without assigned representation is possible. The logical model is, however, specific instances in the domain model types. Each object (node) is therefore assigned to its type and its attribute values. Objects, attributes and types of objects are on clearly identified through their names. The graphical model is separated from the logical model and provides topological structures and abstract objects. Not all of these objects have a visual representation. There are also visual objects without logical or topological mapping function, such as building data. Each node graphic is optional, and therefore an object of the logical model. Thus, even indirectly, the presentation of the object is defined. Alternatively, the presentation can also be directly assigned, for example, to a building or channel data lines.

![Figure 3: generic data model](image)

4. APPLICATION EXPERIENCE

The continuous support for the lifecycle process of a building in the developed software framework has been applied in two different fields of automation applications. It showed up in the framework through a cooperation with a large manufacturer of cleanroom systems, that in all phases of the plant lifecycle data produced consistently and without media breaks can be managed. In addition, a so-called energy-control system on the basis of the framework in collaboration with a maker of monitoring and control software for renewable energy is created. This system transferred the configuration from a single 2D drawing file systems for photovoltaic, wind and hydro power plants and offers a generic way to visualize the individual components and to control it via the Web (cp. figure 5).
The complex problem of cleanroom control, as one of the two applications, it is clear from the large number of fans (known as Fan Filter Units), typically 500 to 25,000 units, which are depending on the associated space expiring production in different work arrangements air with reduced proportions of particles blow. The monitoring and control systems regulate the smooth and energy-efficient operation of the whole system, which is a common communication and interaction with other building automation systems. For example, the event of a fire hazard, the corresponding signals from the fire protection system are received and interpreted accordingly, so that geographically nearby fan to prevent the fire from spreading. The planning of such plants will be conducted with 2D drawing programs, where the modelling of components and compounds in developing specific guidelines set primitives. In the current industrial practice, this information may be for the installation and repair not automatically be accepted. The physical link between the components must be manually matched with the planned configuration and should be examined for compliance. For the configuration of the interaction with other building automation systems, the spatial data in turn has to be done by hand (cp. figure 4).

A special adaptation framework, the expectation for continuity of data along the software system is fully met. All relevant components and compounds were from 2D AutoCAD drawings with preconfigured interpreting scripts provided. So were the object models for individual software systems throughout the plant lifecycle completely automates generate. A redundant manual data entry for the commissioning, monitoring and control system was therefore completely avoided. The examples are configured clean rooms were available in any internet visualized and necessary control functions via Web services are running.

![Image of graphical user interface](image.png)

Figure 4: The graphical user interface of the domain configuration of the clean rooms control system

The second scope, concentrates on the advantages of model transformations for the configuration of monitoring systems and the benefits of a harmonized view of heterogeneous systems with data coming to the domain of energy from renewable sources plants. Optimum energy yields from such investments are created in energy parks, with the operators of electricity from different energy sources, such as photovoltaics, wind power and hydropower combined. In order for them to act quickly will be a simultaneous viewing of relevant information from all sub-systems is essential. The heterogeneity of the systems and types of plants makes an integration of the software systems virtually impossible. Furthermore, the Energy Park for the whole configuration information is only created manually or with complex direct interfaces with proprietary systems, so that future changes opaque and are difficult to be carried out.

The framework based Energy Control System (ECS) has a solution to the demonstrated problem. The planning of the energy park was seen as 2D models drawing, resulting in the configuration data for all relevant software systems of the plant life along with the corresponding interpretation script. As an example in figure 5 spatial information

- for a Google Maps web realized overview,
- for a management system for asset master data and
- for a monitoring and control system for photovoltaic installations is demonstrated.
All existing data in the system could be visualized on a web-based interface. Through this for the first time the data of a wind and hydro power plant without additional implementations (only configuration) in a harmonized view are presented side by side.

Figure 5: Application examples in the domains of renewable energy systems (schematic representation)

4. SUMMARY AND OUTLOOK

Building automation is becoming increasingly important both in the industrial and in the private sector, for example, by providing flexibility in the support of old / disabled people or contributing to everyday comforts. But also because of the multitude of possibilities for energy-efficient buildings in the operation of private and business buildings the importance of such systems is steadily increasing. Commissioning or changes in the buildings require the transfer of spatial configuration of automation components in several heterogeneous tools, often with a huge manual effort and constantly repeating routine activities. With this present contribution, the concept and application experiences of the FZI Forschungszentrum Informatik Karlsruhe (Germany) a newly developed software framework has been presented. This allows the consistent and lifecycle-oriented management of complex spatial configurations of automation systems in buildings. It also allows small and medium-sized components, system vendors and system integrators to offer modules that help to easily integrate powerful and individual IT solutions along the building automation life cycle.

The objective of future development of the framework, is to put focus on the energy sector. It is planned to use a user-specific configuration and modeling of the energy supply, in various detail and quality levels, evaluations of energy depending on the usage behavior. For this appropriate presentation and analysis options have to be available, based on a unique space-time event-based relationship. On one hand users get an accurate overview of their specific energy consumption behavior in the various phases of use, such as holidays, work week or weekend and are thus optimized in terms of energy-aware behavior. On the other hand, the automated derivation of individual actions and identification of weaknesses in the configuration of supply and provision of user behavior, can result in significant energy savings.
REFERENCES


Improving Process Performance through SMEs-SPC

Mohd Nizam Ab Rahman1*, Rosmaizura Mohd Zain1, Zulkifli Mohd Nopiah1, Ahmad Rasdan Ismail1, Jaharah A.Ghani1, Baba Md Deros1, Norhamidi Muhamad1 & Suriani Abdul Rahman2

1Department of Mechanical and Materials Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia 43600 Bangi, Selangor, Malaysia
2Faculty of Information Technology & Quantitative Science, Universiti Teknologi MARA (UiTM). Bukit Ilmu, 18500 Machang,. Kelantan, Malaysia

ABSTRACT

In the present era of globalization, SMEs need additional support to stimulate forthcoming business opportunity. It totally depends on the best strategies to improve quality performance and reduce the defect rate of manufactured products. This improvement could be achieved through the practices of SPC to control, monitor and managing process specifically at production site. SPC can be conducted as simple as analyzing data and plotting charts. This paper highlights the outcomes of an effort to develop the SPC computer based system intentionally to perform simple statistical analysis as well as to manage the quality data processes. The system named as SMEs-SPC (Small and Medium Enterprises - Statistical Process Control), has tended to focus on particular data sets, simple statistical operations and users group. This paper is a continuation study which focused on the development and application of SPC system in a case study company. The results showed that the SMEs-SPC applications are applicable to all types of processes, machines and parameters to store the information, data and subsequently generate the graph. The utilization of this system may help organization to train and educate employees on SPC application, which could become one of the strategies to improve the quality management tool. The value of this study would be particularly useful to SME companies that are contemplating the adoption of advance manufacturing technology.

1. INTRODUCTION

According to O’Regan and Ghobadian [1], strategy is commonly to have a few different numbers of meanings. A definition by McDonald [2], strategy encompassed some elements such as focus on long-term performance of the organization, determine the activities of the business that parallel to the environment toward stimulate maximum opportunities and propose the organization’s activities based on resources available. Indeed, deploying strategy enable the company to expand the business as efficiently as possible towards the performance improvement. Samat et al. [3] believe most organizations in Malaysia have started to concern about quality as an essential part of their strategies in order to meet challenges and improve their positions in the marketplace.

Nowadays, Statistical Process Control (SPC) method has been implemented with success in many large organizations but there is still less efficiently used in smaller organizations specifically in Malaysian Small and Medium Enterprises (SMEs) manufacturing companies. However, Antony et al. [4] believe due to increasing of high capable business and competition in domestic and global environment, large organizations mostly depend on SMEs to deliver higher quality of products at lower cost. Thus, SPC has evolved into a business strategy in many large organizations. Therefore, SMEs has left no choice other than to consider the introduction of SPC methods at their production site.

This paper aimed to report on the new dynamic system (as a tool) for manufacturers in SMEs in order to perform simple statistical analysis and manage quality data processes. The system named as SMEs-SPC (Small and Medium Enterprises - Statistical Process Control), has tended to focus on particular data sets, simple statistical operations and users group. In addition, this study attempts to assist SMEs to make quick corrective action or instant decision making of quality problem, as well as encourage production workers from operator level to adapt with the SPC computer based efficiently rather than traditional SPC or manual techniques. Dangayach and Deshmukh [5] contend that, the effort to reduce operating costs and improved manufacturing efficiency has realized some
manufacturing organizations to further experience on direct and indirect Advanced Manufacturing Technology (AMT), or can be described as group of computer based technologies, including Computer Aided SPC (CASPC) or SPC computer based applications. Therefore, SMEs-SPC system described in this paper will contribute and offer some further advantages.

2. THE NEED FOR SPC IN SMEs

Generally, the long-term goals for SMEs are to persist in business competitiveness, survive, raise and make profits. One of the fundamental ways to achieve this goal, SMEs are urged and must reduce several items of failure costs or waste items including rework, scrap, warranty claims, product recall and others. In their studies, Ab. Rahman et al. [6] believe SMEs plays an important role in gaining competitive advantage. According to them, there are two fundamental reasons to use SPC in SMEs; SPC is necessary because large companies will require high quality product at minimum possible costs and it is the survival necessity in this high competitive business environment. According to Xie and Goh [7], manufacturing sectors have benefited from the implementation of SPC tools/methods to lead in decision-making process. In essence, control charts and process capability utilized through the implementation of SPC are proposed to identify and enhance the process performance. This fact also supported by Surak [8], suggested that there are several resources to implement SPC, such as software selection, control chart, histogram and capability indices to identify the improvement items in process. In addition, control chart is useful because it able to distinguish between assignable and common causes of variation in the process [9], while process capability purposely to measure the actual output of a process and then compare the capability to the specifications given for the process or product involved [10; 11]. Alternatively, process capability consists of three primary techniques including histogram or probability plots, control chart, and designed experiments [9].

According to Antony et al. [4], the utilization of CASPC has received more attention by many organizations. For example, many companies preferred to use spread sheet software to develop control charts. However, many manufacturing organizations still remain on using traditional SPC or manual practices (e.g. paper-based control charts). Consequently, traditional SPC focused on particular limitations where only little quality faults are detectable, time-consuming and burdensome. Moreover, they do not really understand the purpose of quality data collection and there is no systematic way for them to efficiently manage and organize quality data processes. Furthermore, Hofmann et al. [12] stated that a problem regards to statistical software in company is not portable and multiple users since the system developed for one specific computing. From researchers’ point of view, quality data or information could be shared via online or server application to perform data analysis, as well as to provide reports in the standardized formats. The researchers have considered this limitation in developing the system.

From the previous study [6], the researchers have typically identified the problems of using SPC and the barriers that may prevent the SMEs in Malaysia from adopting SPC software or system effectively. The results revealed that some problems (found in the preliminary study), for example lack of efficiency of quality data management, calculation errors, inconsistent format, limited access in data collection, complicated specifically for operators and the system or SPC software is very costly, while some barriers are also exist due to the lack of awareness to the importance of SPC applications. These facts indicate that the implementation of SPC is believed to be appeared at a moderate stage in Malaysian SMEs. Lack of top management support, high cost of technology and lack of skill workers are also the barriers to SPC implementation. Therefore, this paper will focus on the development of SPC system computer-based which the system offers at lower cost, simple statistical operations, ease-of-use for production levels, extensible to access data at flexible times; reduce time to input and retrieval historical data, prevent errors in calculation, trigger out-of-control data automatically, avoid late/delay information on out-of-control event in production, provides interesting features as well as standardized report and consistent format. Therefore, in this research, with the co-operation between government, industry and university, it outlines the development of SPC computer-based system for conducting statistical analysis on production processes as well as, assists the company operations performances.

3. RESEARCH METHOD

A case study can be used when attempting to understand complex organisation behaviours. The case study is especially appropriate when trying to answer the “how” and “why” questions in exploration of the research problems. The case study typically involves a small number of cases which are not necessarily represented of the larger population. Data collection for case studies may come from various sources such as interviews, company documents, archival records, observation and physical artifacts. Moreover, this method need not always include
Improving Process Performance through SMEs-SPC

direct and detailed examinations as a source of evidence as required in qualitative research [13]. To meet the aim and objective, primary case studies were conducted at ten manufacturing companies (SMEs) located around Selangor (Klang Valley). In this period of study, data and information were gathered from various sources such as interviews, survey (questionnaires) and observation. The type of companies involved includes automotive, electronics, medical disposal devices, computer components, plastics and chemicals. Most of the engineers or managers involved were from production or quality department, since this study focused on the production site. In general, this study is concerning on company background, perception and level of SPC practices, types and characteristics of SPC software used, advantages, barriers, suggestions and, seeking for potential companies to develop/use the SPC system.

The study is enhanced and involved case study investigation and collaboration with the automotive company, by focusing on process characteristics, collection of existing quality data processes, process control sheets (paper-based control charts) and users group involved in the process. The study encompassed design and development of SPC system based on process control sheets and control process plan provided by the company, as well as consulted by company requirements and academician suggestions. However, this paper is only focus on the first phase development of SPC system which involved initial trial or pre-testing of the SPC application based on existing data provided by the company. Future study will focus on the development of online SMEs-SPC system based on local area network that provided the means to share software and data information in solving quality problems at the case study company, and other SME types of business/operation that compatible with the SMEs-SPC system.

4. RESULT AND DISCUSSION

The studies’ outcomes are summarized based on the simple SMEs-SPC application, developed for a case study company, named by company A which is simple and not too complex as compared to other commercial statistical software. Note that, the system may be used effectively by anyone who does not have a strong background in statistics specifically for operators at process/production sites. In short, SMEs-SPC system is integrated by client server database that refers to MySQL (Structured Query Language) with back-end relational user interface developed in Microsoft Visual Basic. All tools or object models of visual basic are used to access SQL server. Based on stored data in database, visual basic is responsible to compute or performing statistical calculations. User interface provides the user with intuitive graphical tool to analyses data through control charts and histogram. According to this, the MySQL is supported by admin software to manage a whole MySQL server such as browse and drop databases, create tables, views, rename, export and import databases and others as shown in Figure 1. On the other hand, developing SPC system through a main server application allows many users to use the same software, share the same data sets and sharing knowledge about capabilities, limits, graphs and out of control event. For example, these three contexts in company A encompassed production to quality control (QC) to top management in order to inform and make instant decision making of quality problem in process site. Figure 1 shows the schematic diagram designed for company A which involved MySQL server database and admin application that assigned to manage database as well as directly linked to the user workstations. Currently data quality processes or variable data is not only to monitor the variation of process, thus it is used for the presentation of standard reports and graphs. It can facilitate end users to compare process stability performance across different time, values and process.

![Figure 1: Schematic diagrams for SMEs-SPC connection system](image-url)
All users must confirm their identification with MySQL admin to enter onto this system by login the username, passwords and company name. The results will pass directly to admin before user allowable to enter the SPC system. There are some privileges provided by this system such as multiple users or easy to access and enable to use at flexible time around local company; enable to trace the user who delete or edit data values; enable to protect confidential data by security password given. Detail of how it works or data flow diagram is shown in Figure 3. This section is responsible for generate stored data. MySQL admin is typically consisting of six separated tables in order to differentiate data values and add variable information, as well as to assist in computing statistical operations. SPC data processing takes the user’s information or parameters to be applied such as variables used, types of control chart to perform, subgroup size, time and date entered in data entry as well as user define control limits to manage according to the table/report respectively.

![Figure 2: Data flow diagram](image)

In this section, SPC data processing is responsible to organize the stored data and finally passes these results into SPC information system for graph generation and process capability underlying of statistical formula. Thus, development of this system tended to use more graphical common button that offers interesting key features rather than spread sheet software. It is easier to allow user to control the button and generate calculation, graph and report by following simple instructions. For example, users enable to define their own control limit and select control chart rules suit to the company requirement. In some cases, the system is facilitated with code color either red or yellow to trigger out of control event. Feedback action via email can be executed to inform this event. This part becomes very important to inform real time problem occurs in process or production.

From the researcher’s studies, the calculation or SPC operations required by SMEs companies are basic enough such control charts (X bar-R charts and X bar-S charts), histogram and process capability indices. One important issue is regarding on calculation accuracy, as compared to established or commercial statistical SPC software/system, is that the statisticians have verified its calculations. To address the issue, the calculation results in SMEs-SPC system were checked against Excel (programmed of SPC formula) and SPSS. Finally, compare the existing process control sheet or data provided by company A to verify their accuracy with this system. The only case in which the results did not agree was in the calculation of normal curve as well as it shapes and varies slightly.
between established packages such as SPSS and Excel. SMEs-SPC system used the formula of normal curve proposed by Smith [14].

In the context of this system, it development is not intended to be a replacement for the complete and complex statistical operations offered by other/current commercial software packages, which currently used by industry (such as Minitab, SPC Excel and Microsoft Excel). Another important feature, this system is not based on spread sheet or in tabular form and also does not require function or syntax name for conducting statistical calculation as compared to the existing commercial software. Finally, the instruction given in 'help' button is user friendly and easy to understand by all levels. Other important features, the results of SPC data processing by SQL queries are placed into a result table with a standard format and label, to enable processing by any components of the system. For example, the result table stored the main process folder, sample size and subgroup size, and running data including specifications, target, data values and others (e.g. see Figure 3). Another enhancement in this system is the possibility to create final reports that consists of image files and text results. In any case, the possibility of creating presentation-quality reports via SMEs-SPC system is an appealing one.

![Figure 3: Running data](image)

4.1 APPLICATION AND USER INTERFACE FEATURES IN SMEs-SPC SYSTEM

In order for SMEs-SPC to be useable specifically by production levels, graphic interface provided in this system is one of the advantages that attract the users’ interest in performing systematic data calculations. At initial method of application, users should input raw data or quality data process into the system to perform statistical analysis quickly and easily. Then the result displays either as text or in image form. These results should be able to identify whether the variation occurs in control or out of control. The system also performs capabilities calculations to determine the capable of process.

![Figure 4: Login information](image)  ![Figure 5: SMEs-SPC system introduction](image)

In addition, SMEs-SPC is aimed to suggest improvement of some problems as found in the preliminary study. Specifically to enhance the efficiency of data quality management, encourage operators or productions level to handle SPC based on computer rather than traditional SPC (paper-based control charts or manual calculation) and
The following figures address interfaces features available to the user and it is expressed in each interface. A beginner of user is requested to login information, as shown in Figure 4. Thus, this way is possible for protecting confidential data. As mentioned earlier, this system should be used by a wide range of operator levels. Then the user will be given simple description regarding on introduction and key features; objective and application of X-bar & R charts, X-bar & S charts and Histogram, as well as guidelines for using SMEs-SPC to perform simple analysis. It also does have several intriguing features to provide guidelines or information aided by ‘help’ button as shown in Figure 5. These guidelines make use of few key concepts in statistics, which are important to understand when using the system.

The creation of main list or recorded data set in Figure 6 is possible to view all store information. This part allows user to retrieve previous data, modify, continue add and delete data set parameters including sample size, subgroup size, specification limits, date and others. Moreover, the user is also allowed to quickly adding, deleting and renaming main and part folder used in his or her works.

After performing the main and part folder, add variable interface as shown in Figure 6 enable user to fill up either in wording or numeric text regards to subgroup characteristics and control chart setting. Shown in Figure 7 is an interface data entry according to sub-group size (n). Once (n) is set at three in add variable, empty table with three rows are visible automatically to facilitate user to key in data.

The system can perform a set of data table or quality data processes collected at particular day per month. This interface contains update function to delete or edit data. User can express perform correction of value errors in this
part. Control chart regulation list in Figure 8 aimed to reflect the graph output in relation to specifications. These rules can be adjusted to suite the company requirements. Control charts with subgroup size \( (n) \) in Figure 9 will trigger in red or yellow point if abnormal or out of limits occurs. It should be noted that, there are two options of graph to generate which consists of average and range charts \((X \text{ bar } & \text{ R})\) as well as average and standard deviation charts \((X \text{ bar } & \text{ S})\). Control chart shows the variation of values underlying on y-axis versus X-axis. The labels on X axis are sorted either by date, subgroup size or frequency.

Additionally, user are enabled to report any event if control chart rules are violated, through email as shown in Figure 10. This enhancement allow user to create any comment or feedback action associated with the process. This report can include control chart file image, which could be attached in email. The histogram in Figure 11 reports the frequency of measurable data from the process problem. It also indicates the capability of process, shows whether data is falling inside the bell-shaped curve or within specifications as well as target values achievement.

In the final stage of this system, an overall performance report will be generated and it includes control charts, histogram and capability study as viewed in Figure 12. It can be saved as an electronic file by selecting the destination of file and print for company reference.
In addition, the report of process/production performance is presented in a similar or consistent format, which allows the user to show the progress report to customers periodically. It should be noted that, this interface application is only applicable for SMEs-SPC system, to facilitate user in conducting simple statistical analysis. However, there are several parts of SMEs-SPC which could be developed further in order to meet the needs of new user in SME companies or to expand the user base as well as to make it robust.

5. CONCLUSION

In this paper, the SMEs-SPC system development has shown several factors could influence usability of this system from production level. These graphical presentation or user interfaces shown in figures (above) are simple enough and ease of use for them to conduct simple and quick analysis. Another factor, this system attempt to approach SME companies to manage or organize quality data set as efficient as possible, which based on database admin and graphic interface application. This system is applicable to all types of processes, machines and parameters to store the information and data set and subsequently generate the graph. The internal system has demonstrated a data flow mechanism to separate six types of database from interface. Entirely, user is not required to classify all data or information manually. Therefore, in order to remain competitive in business environment, SMEs must continuously adapt to the rapidly changing business strategy. One way is to improve their business practice through the use of latest technology.

As the conclusion, the researchers believe that the utilization of this system may help organization to train and educate their employees on SPC applications, which could become one of the strategies to improve the quality management tools. Future study will concentrate more on the trial runs at the company A and perhaps, in other SME companies, in order to strengthen the application of SPC system and to verify the effectiveness of SPC computer based system rather than traditional SPC practices.

REFERENCES

**An Industrial Validation of Artificial Vision Techniques for the Control of Future High Speed Forming Processes**

P. Fillatreau\(^1\), N. Arana\(^2\)*, E. Saenz de Argandoña\(^2\), A. Izagirre\(^2\), I. Zuriarrain\(^2\) and R. Pop\(^3\)

\(^1\)Delta Technologies Sud Ouest
2 Impasse M. Labrousse – BP 1137
31036 Toulouse, France

\(^2\)MGEP
Mondragon University
Mondragon, 20500, Spain

\(^3\)Institut für Umformtechnik
Universität Stuttgart
Stuttgart, 70174, Germany

**ABSTRACT**

Forming processes are manufacturing processes based on deformation of raw material applying pressure in one or several stages until getting the final product. Correct forming of parts is very important as any defective part may result in big economical losses, e.g. the return of a complete set of parts or the loss of some clients. Real time human supervision of such high throughput processes is not possible. These processes need to be supervised and controlled by sensor based systems allowing to achieve zero-default forming processes, while minimizing costs and maximizing the throughput of parts. On the other hand, commonly integrated sensors (force and acoustic) do not allow detecting all types of defaults for this kind of applications (i.e. local burrs due to punch micro cracks). Thus, in the framework of our European Craft Pro2Control project, carried out by leading German, French, Italian and Spanish companies, universities and forming industries, we have considered including an artificial vision (AV) system. An AV system may allow better understanding and/or characterization of force and acoustic measurements, while it can detect defaults that the other sensors may not. While acoustic or force sensors provide real-time information, Vision systems are often too slow for the targeted parts throughputs. Because of that, for the AV part of our control system, a high speed FPGA-based smart camera is proposed. This paper presents the smart camera architecture, as well as the methodology of the image processing algorithms selection and implementation on its internal FPGA (electronic real-time processing device based on reconfigurable logics). The needed image processing algorithms have been analyzed, so that the most time consuming ones (early image processing steps, usually parallelizable) are implemented in the FPGA of the smart camera. Those algorithms are the camera geometric distortions correction, image thresholding and (optional) compression, so that corrected binary images (optionally compressed) are delivered through the internal USB connection to a standard PC, where high level less time consuming algorithms complete the image analysis processes. We also present the architecture of the complete multisensor system. And finally, we present the results obtained for the control of a forming production line in a real industrial environment producing mechanical parts for the automotive industry.

**1. INTRODUCTION**

The manufacturing industry plays a vital economic role all over the world. As an example, manufacturing activity in Europe represents approximately 22% of the EU GNP [1] and manufacturing-related activities account for around 75% of the EU GDP [2]. The current market scenario, with the emergence of new developing countries in Asia, is pushing sheet metal component manufacturers to be more and more competitive concerning their production costs. This is mainly due to the fact that the automotive industry is demanding increasingly complex parts, manufactured with lower formability materials, using simpler production processes at higher rates, with stricter quality requirements. As a result, even small variations in the properties of the raw materials or in the process variables can produce low quality parts.

However, good forming of the parts is a very critical factor, as any defective part may result in big economical losses, e.g. the return of a complete set of parts or the loss of some clients.

* Corresponding author: Tel.: +(34) 943 739640; Fax: +(34) 943 791536; E-mail: narana@eps.mondragon.edu
2. CONTEXT

Pro2Control is a European Craft project formed by a consortium of leading German, French, Italian and Spanish companies, universities, research institutions and forming industries, with the aim of implementing a zero-defect forming control system, trying to minimize development costs and at the same time maximizing the throughput of parts. In order to build a zero-defect control system, an “intelligent control system” has been developed. This control system is based on Knowledge Based techniques (in this case a rule based Expert System), and gives the operator of the machine feedback and advice about the possible errors/mistakes, and what to do in order to correct/prevent them.

This system is fed by data provided by acoustic-force sensors and by an AV system [3]. On one hand, the acoustic-force sensors are useful to detect defects due to mechanical faults in the press (e.g. tooling breakage, bad evacuated parts or material scraps inside the tooling). On the other hand, AV systems are useful to catch undetected defective parts (e.g. localized burrs or geometrical defects and dimensions outside required tolerances).

The industrial end-user partners have defined which parts to control, as well as the maximum processing speed. The two references presented in Figure 1 correspond to two representative flat parts studied at the present research. Their production throughput is about 60 strokes per minute, and two parts are produced per stroke. One of the aims of the Pro2Control project is also to design a control system able to cope with the high throughputs of future presses (about 1800 parts per minute).

![Figure 1: References studied at the present research](image)

In our proposed AV system, an FPGA-based programmable intelligent camera (Figure 2) with a USB 2.0 connection has been built. The bottleneck algorithms are implemented in this camera, and the on-board processing results sent to the PC. In the rest of the paper, this mixed architecture, as well as the developed algorithms and obtained results are presented.

2.1. OBJECTIVES OF THE AV SYSTEM

The main purpose of the AV system is to evaluate the quality of the produced parts, reject the defective parts and continuously inform the intelligent control system about the quality of the parts. The defects that can be detected by the AV system can be divided into two main groups: those visible from above, e.g. dimensions outside tolerances; and those visible laterally, e.g. non-planar parts, parts with localized burrs and thickness outside tolerance. Therefore, the
AV system involves two iCam cameras that take one picture each (upper and lateral views) of every controlled part [4].

The complete AV system is composed of two “iCam” cameras and a PC. The AV system is mounted on a ramp, where the parts fall after leaving the press. The images are processed and, depending on the quality of the part, a sorting device rejects the low quality parts (Figure 3).

3.- DESIGNING AND VALIDATING THE IMAGE PROCESSING ALGORITHMS IN A PC

Before implementing the image processing algorithms in the “iCam” FPGA, we have designed and validated them on a PC. This way, we can detect the bottleneck algorithms, i.e. the most time consuming ones.

As the AV system has to analyze the geometric dimensions with high accuracy, the first step is the geometric distortions correction. We have extracted the internal camera parameters by observing a chess grid pattern from multiple views, using Zhang's method [5], and OpenCV *ad-hoc* libraries. This process is made off-line, and provides a distortion map. Furthermore, a program has been implemented that allows to extract the external parameters of the camera, using a novel method [6], so that the upper view camera can be positioned parallel to the observed surface, minimizing perspective deformations.

After the geometric distortions have been corrected, the image is thresholded and all contours are detected and filtered depending on their size, in order to detect the part contour without being sensitive to image noise. We calculate the bounding box of the part, and extract the part’s external circular contour points, in order to find the centre of the part. A least squares circle fitting of these contour points is performed in two steps. The first one computes an initial estimation of the centre, and the second step deletes points that do not belong to the fitting circle, allowing to refine its estimation (Figure 4). Finally, the dimensional features to be controlled are extracted using image analysis algorithms and the pixel to mm transforms provided by the calibration step [5,6].

Figure 5 shows the different image processing algorithms implemented. They can be decomposed into 3 steps: 1) part detection, filtering and distortion correction, 2) part centre detection, and 3) dimensions calculation.

The image processing algorithms have been implemented and tested for both selected parts types (see Figure 1) on a PC (Pentium 4 at 3Ghz) using OpenCV libraries (for image processing), and Gtk libraries (for Graphic User Interfaces). The processing time of each algorithm has also been evaluated.
Table 1: Timing information of the AV algorithms

<table>
<thead>
<tr>
<th>Part Detection + Image filtering + Distortion Correction (Step 1)</th>
<th>Ref.1</th>
<th>Ref.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33 msec.</td>
<td>1818 parts/min.</td>
</tr>
<tr>
<td>Center Calculation (Step 2)</td>
<td>0.64 msec.</td>
<td>93720 parts/min.</td>
</tr>
<tr>
<td>Dimension Calculation (Step 3)</td>
<td>4.4 msec.</td>
<td>13440 parts/min.</td>
</tr>
<tr>
<td>Total time Calculation</td>
<td>38 msec.</td>
<td>1580 parts/min.</td>
</tr>
</tbody>
</table>

4. VISION SYSTEM: A HARDWARE/SOFTWARE CO-DESIGN ARCHITECTURE APPROACH

FPGAs are programmable electronic targets based on logic elementary blocks. The main features provided by FPGA-based implementation of data, signal or image processing (when possible, opposed to microprocessors or DSPs) are the following [17]:

- Processing times are reduced by a factor 10 with respect to an implementation on DSP, and 100 if compared to an implementation on a microprocessor [7].
- FPGAs allow to fit the hardware implementation architecture and the structure of the implemented algorithm (algorithm architecture adequacy to optimize the implementation).
- Architectures based on FPGAs permit higher capacity of miniaturization.
- FPGAs architectures are re-programmable.
- FPGA-based architectures provide higher confidentiality protection.

Implementing advanced algorithms on FPGAs is not an easy task, compared to the same implementation on regular PCs, as it requires advanced skills in the fields of both image processing and electronics. The language selected to implement the algorithms is VHDL. It is an advanced implementation language intended to describe hardware architectures. The design of a proper implementation needs to make a thorough electronics design and is generally performed under strong development constraints given by the used hardware architecture and its resources (available logical elements and memory) or by the targeted application (processing times required, algorithm complexity, operands precision). To cope with them, solutions like parallel processes implementation, “pipeline” architectures suited to the input data flow, algorithm architecture optimization, implementation of hierarchic sequencing architectures, can be involved.

![Figure 6: “iCam” architecture](image)

While smart cameras based on standard embedded microprocessors (e.g. PowerPC) or DSPs are now available commercially and allow reaching a new range of real time performances [8], finding smart cameras based on FPGAs or IPs in VHDL for advanced image processing is more difficult. Standard in-camera processing functions for commercially available FPGA-based cameras are: sensor configuration (e.g. gain, offset), Region of Interest programming, image subsampling, thresholding or image convolution. Many research teams are facing the problem of implementation of advanced image processing algorithms on FPGA or developing FPGA-based smart cameras [9,10],
defining direct co-design tools for the rapid prototyping and the implementation of distributed real time applications on mono or multicomponent architectures [11,12], or defining direct generic VHDL synthesis tools [13].

Delta Technologies has been facing image processing applications demanding extremely high processing rates (seeds production control of 1000 sunflower seeds per second, image processing on satellites at the acquisition rate of the future hyperspectral satellites, high speed obstacle detection for vehicles using dense stereovision required to run at 100Hz [14,15,16]). To solve these problems, the corresponding image processing algorithms architectures have been adapted for implementation on camera’s FPGA and several smart FPGA-based cameras have been built. Figure 6 presents the current architecture of the “iCam” camera.

We aim at coping with the throughputs of future presses and want to reach a potential throughput of 1800 part per second. From the results presented in Table 1, we can conclude that the bottleneck, corresponds to Step 1 algorithms, i.e. the ones corresponding to the geometric distortion correction and noise filtering. The corresponding algorithms are highly parallelizable and thus should be implemented in the camera’s FPGA. Step 2 algorithms processing times are small and compatible with the targeted processing times. Thus, although they are highly parallelizable, they are run in the PC. Step 3 processing times are also compatible with the throughputs we aim at. The corresponding algorithms' standard architectures are not well suited for FPGA implementation, because of their complex decisional structure, so they are implemented in the PC at this stage of our work.

5.- IMPLEMENTATION OF THE IMAGE PROCESSING ALGORITHMS ON FPGA

The used CMOS sensor provides 10 bits data. We have implemented the following algorithms on the camera’s FPGA:

- **Distortions correction**: when the correction of distortions is activated, we reduce the pixel resolution down to 8 bits due to constraints on the pixels memory buffering involved.

- **Image thresholding**: the output is a binary image (1 bit/pixel). Implementation of this process is straightforward and thus it will not be addressed in this paper.

- **Binary image compression (Optional).**

5.1. - CORRECTION OF THE “ICAM” DISTORTION

On-board the camera, we consider the distortion correction in pixel coordinates, in order to avoid transforming pixels into real world coordinates (i.e. when the image plane of the camera is parallel to the observed surface the pixel/world unit rate is constant) [6].

To correct the distortions, several strategies are possible. A simple one consists in saying that the corrected and distorted image have the same size (NbLin, NbCol) and that the pixels co-ordinates values can only be included in [0, NbLin-1] and [0, NbCol-1]. This may lead to (small) changes of the field of view, but simplifies the algorithms by avoiding to manage image mapping / size change algorithms.

For our pipeline on-board implementation, two approaches where possible (see figure 9):
• Use transform T to compute the corrected pixel position for each incoming pixel of the input distorted raw image. This approach seems the most efficient from a real time point of view, as it would allow to compute the transform as soon as a single input image pixel is available. This approach has two major drawbacks: (1) it implies inverting the classical distortions equations, and (2) it may lead to “holes” in the resulting corrected image (the process does not guarantee that every pixel in the corrected image is computed), or even to distinct pixels of the corrected image to project on the same pixel of the distorted (raw input) image.

• Thus, we decided to use transform T’ instead, in an algorithm where each pixel position in the corrected non distorted image is considered, and we compute the corresponding pixel position in the raw image. From this position in the raw image, we can then compute the grey level of the considered pixel in the corrected image by considering the grey levels in the neighbourhood of its projection in the raw image. We used as typical neighbourhoods elementary cells of $2^N$ pixels, with $N=2, 3$ or $4$.

![Figure 8: Radial distortions correction](image)
![Figure 9: Image transformation](image)

To compute the positions of the corrected non-distorted pixels positions, we encoded the polynomial corrections into distortions correction tables, downloaded via USB to the “iCam” camera. These tables have to be stored in non volatile memory, i.e. Flash memory. But Flash memory has a limited bandwidth (in our case, 10Mbytes/s while the camera sensor frequency is 40MHz) and depending on the required real time performances, it may be necessary to transfer the tables to memory with a better bandwidth (external RAM on-board the camera in our case, with bandwidth around 400 Mbytes/s).

When computing the value of pixels along a given line of the corrected image, we may need pixels of the distorted (original input) image located in the grey areas (Figure 10). In our pipeline computation implementation, we manage a constant input image buffer size; to define this (maximum) size, we need to know what the maximum amplitude (in number of lines) of the distortion is. We also had to discuss which kind of available memory we should use for the real time storage of the pipeline incoming image pixels buffer (Figure 10). Embedded M4K in our FPGA (52 embedded buffers of 512 bytes each) allow parallel access to all memory elements and makes the design simpler (the design is more complex if you use external RAM for example), but their limited number sets an upper limit to the maximum distortion (and thus incoming image pixels buffer) the system can cope with.

![Figure 10: Maximum distortion size defines the size of original image buffer in the pipeline process](image)
5.2. **Image Compression**

This step processes binary images. We have studied several approaches to implement this step on FPGA:

- A first way to implement this compression is to represent the compressed image by chains of numbers of successive pixel of the same level (’0’ or ’1’). We need to be compatible with the maximum case of an image where all pixels have the same level.

- Using a similar approach but compressing the image line by line is more efficient. For each line, we need an additional signal (“end of line”).

We also have to represent the level (’0’ or ’1’) of the successive pixels. We could think of adding a bit to each number encoding the successive pixels having the same level. A more efficient way is to choose a default initial level and to use a bit flag to represent the two possible cases (initial image pixel level is the default level or not).

Those two approaches are not very efficient as they involve the manipulation of large integers (encoded on a large number of bits). Instead, we have decided to manage 8 bits data, and, as we also need to encode at which level (’0’ or ’1’), the successive pixels are, we use one bit (say, the first one) to encode this information. If the number of successive pixels having the same level exceeds $2^7-1=127$, then we will have several successive 8 bits numbers encoding pixels having the same level (’0’ or ’1’), encoded by the first bit.

6.- **Results**

The “iCam” camera has been built and tested in an AV system (see Figure 3). As explained earlier, there are two “iCam” cameras in this prototype. The first one (upper view) measures dimensions A, B, C and D (see Figure 1); the second one (lateral view) measures the localized burrs.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Lower tolerance (mm)</th>
<th>Upper tolerance (mm)</th>
<th>Average (mm)</th>
<th>Standard Deviation (mm)</th>
<th>False negatives (%)</th>
<th>False positives (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension A</td>
<td>6.10</td>
<td>5.8392</td>
<td>5.8392</td>
<td>0.0087</td>
<td>0.6089</td>
<td>0</td>
</tr>
<tr>
<td>Dimension B</td>
<td>6.10</td>
<td>5.8706</td>
<td>5.8706</td>
<td>0.0095</td>
<td>2.4046</td>
<td>0</td>
</tr>
<tr>
<td>Dimension C</td>
<td>2.55</td>
<td>6.10</td>
<td>2.6987</td>
<td>0.0033</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dimension D</td>
<td>33.75</td>
<td>34.5</td>
<td>34.0849</td>
<td>0.02373</td>
<td>3.9600</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2 summarises the results achieved with the upper camera during the testing period in a company. The table specifies, out of 3120 measurements, for each measured part dimension, the lower and upper tolerances, the average value of each dimension, the standard deviation of each dimension and the percentage of “false negatives” and “false positives” achieved for each dimensions. The most important values in the table are the standard deviation (that represents the variability of the measurements) and the “false negatives” and “false positives”. The “false negative” items are the number of times that good quality parts were classified as bad parts, and the “false positive” items are the number of times that bad quality parts were classified as good parts (this never happened because no defective parts were produced during the testing period). The presence of local big burrs, due to punch micro cracks, is also now consistently detected by the lateral view camera.

7.- **Conclusion**

This paper deals with the construction of a global control system for metal forming processes, based on Artificial Intelligence techniques. Two types of sensors are used by the control system: force/acoustic monitoring system and an AV system. Details about the AV system are explained, in particular, its mixed co-design architecture based on two smart FPGA-based cameras and a PC.

An intelligent camera (called “iCam”) has been built, where the bottleneck AV algorithms (e.g. thresholding, camera distortion compensation and optional image compression) are implemented on FPGA. The processing steps embedded on-board the camera are executed in real time, at the maximum pixel acquisition rate (40 MHz). Presently,
the rest of the image processing algorithms are implemented on a PC communicating with the “iCam” through a USB 2.0 link. The image processing algorithms on PC only take an extra time of 5 msec.

Finally, the “iCam” camera is fully customizable via a USB proprietary software, allowing to set image acquisition and processing parameters and to recover images and processing results in the PC.

Implemented algorithms as well as obtained results are presented. Our approach provides a fully mastered development, and guarantees durability and maintainability of the system, as well as it maximizes the throughput, making it possible to cope with the cadences of future presses (around 1800 parts/min).

REFERENCES

A Service Target Method for Lead Time Quotation in Assembly Systems

R. Ghasemy Yaghin, S.M.T. Fatemi Ghomi
Department of Industrial Engineering
Amirkabir University of Technology
Tehran, Iran

ABSTRACT

This paper presents a method to assign product due dates for assembly systems with uncertain manufacturing and assembly processing times. The parts arrive to the system as dynamic. The product completion time is the sum of system flow time and arrival time. First, through utilizing the longest path analysis in queueing networks, the distribution function of system flow time is computed by Hartley and Wortham’s method [27] and it is shown that the solutions achieved are significantly much faster than those obtained by Adlakha and Kulkarni’s method. Then the product completion time distribution function is computed by the convolution of system flow time and arrival time. Product due dates are assigned by meeting a service target.

1. INTRODUCTION

Time Based Competition (TBC) has become an important manufacturing strategy in recent years. In an increasingly customer oriented marketplace, low cost and high quality are now considered given and the new competitive dimension is time. Global competition is intensifying the pressure for firms to improve levels of customer service as a source of product differentiation in order to gain a competitive edge [1].

Due to advances in manufacturing systems (e.g. FMS, CAM, CIM, etc.) and introduction of the ideal concepts of inventory control systems (e.g. JIT, ZI, etc.), due-date-based scheduling research has received considerable attention in the last decade and a wealth of literature has been reported in this area. An industrial survey of US companies using FMS indicates that meeting due-dates is the most desirable objective of management [2].

Short and reliable due dates are critical to win customer orders for engineer-to-order and make-to-order companies supplying capital goods [3], [4]. These products are often complex assemblies with many stages of manufacture and assembly.

This paper investigates the problem of assigning product due dates for stochastic assemblies. It defines the due date of an order as the lead time plus its arrival time. In this paper, both of them are assumed stochastic variables. The optimization assigns a product due date to meet a delivery service target. They can be computed quickly and give reliable product due dates.

Due date assignment is often a trade-off between earliness and tardiness costs. For example, a contractual tardiness penalty may be invoked if the product is completed after the due date agreed with the customer; the early completion of activities may result in holding costs. Commonly used cost functions which are minimized in assigning due dates include mean absolute lateness (MAL), mean squared lateness (MSL) and sum of the squares of lateness (SSL) [5], [6]. These cost functions penalize early and late completion symmetrically. Asymmetric earliness and tardiness (AET) may penalize tardiness more than earliness [7]. A small number of researches studied due date management from a profit maximization rather than a cost minimization perspective, considering order acceptance decisions [8]. Another important way of assigning due dates is by specifying the probability of completion before due date; that is the service target ([6], [9], [10], [11].

Some survey papers have been concentrated on specific aspects of due date scheduling problems as, for instance [13], [5]. Some methods use non-stochastic estimates of processing time and neglect assembly operations. In general, the shortcomings of these empirical methods are: (1) they lack efficient methods to calculate the coefficients in their due date expressions; (2) they do not provide information on the distribution of the completion time, which is necessary to analyze risk for a capital project [11].
Analytic methods include queuing networks and linear or non-linear programming. The due date is assigned by minimizing a cost function [13], [9], [14], [6], [7], [2].

Recently, methods and tools based on intelligent system, data mining and information technology have been applied in due date management environment.

The above analytic and new methods were applied to job shops with stochastic processing times. However, these did not include assembly processes. Yano [17] considered stochastic lead-times in a simple two level assembly system, with different processing time distributions including Poisson and Negative binomial. In multi stage assemblies, Song et al. [10] described a recursive method to set stage due dates to meet specified service targets. Song et al. [11] developed an approximate method to assign due dates by minimizing earliness and tardiness costs or to meet a service target. Figure 1 summarizes some of those well-known methods to assign due dates.

Cheng and Gupta [5] and Soroush [2] express that most analytical studies are limited to “small” problems. This paper describes an analytic method to assign product due dates for complex products in the following directions. First, a common assumption found in the majority of the analytical due-date assignment studies is that job processing times are assumed to be deterministic and known before processing begins. However, in many shops, some or all processing times are random variables, manufacturing and assembly processing times are assumed random variables. Second, an implicit hypothesis in the due date based scheduling research is that transit time in buffers is null, i.e., a part which leaves a machine is supposed to be instantaneously available for the next machine (for example [12]) or deterministic (for example [16]), the transport times between the service stations are assumed to be independent random variables. Azaron et al. [19] developed this approach to model assembly systems. Third, in assembly systems completion time is calculated directly and multistage structures are not decomposed into two-stage subsystems. The last issue indicates the solutions achieved by Hartley and Wortham’s method are significantly much faster than those obtained by Adlakha and Kulkarni’s method which also can be useful for real life applications in production scheduling systems.

The remainder of paper is organized as follows. The next section formulates the due date setting problem for multistage assemblies and meeting service target. Section 3 gives a numerical example to demonstrate how the developed method works and efficiency is fully respected for validation studies. Section 4 is computational results. Section 5 is conclusions and recommendations.

![Figure 1: Classification of methods for due date setting](image)

2. PROBLEM FORMULATION

In this section, the completion time distribution function in stochastic assembly systems is computed. Then, by meeting service target, optimal due date is assigned.
Traditionally, material flow analysis in assembly systems has been based on the assumption that the system operates deterministically. In recent years, attention has been directed to a more realistic analysis of assembly systems, explicitly treating the stochastic events that influence operations.

According to procedure, queueing network during a process is transformed to a stochastic one and then lead time distribution function or flow time will be calculated and finally, product due date will be obtained. This algorithm operates sequentially so mean of due date or other moments are calculated. At last through cost minimization of earliness and tardiness and meeting service target the optimal due date will be obtained.

All inter-station buffers are infinite. The time spent in a service station would be equal to the processing time plus waiting in the queue in front of the service station. Therefore, the time spent by a finished product in the system, called manufacturing lead time, would be equal to the length of the longest path of the queueing network whose arc lengths are the transport times between the service stations. The distribution function of the flow time can be approximated by computing the distribution function of longest path in the queueing network.

There are many papers about the longest path analysis in stochastic networks, but not queueing networks. For example, Kulkarni and Adlakha [18] developed an analytical procedure for PERT networks with independent and exponentially distributed activity durations. Fatemi Ghomi and Hashemin [26] generalized the Gaussian quadrature formula to compute longest path function in stochastic networks. Hartley and Wortham [27] developed formulae that may be used to derive the (c.d.f.) "uncrossed" networks. Azaron et al. [19] developed a method for approximating the distribution function of the longest path length in the network of queues by constructing a proper continuous-time markov chain. In this paper, Hartley and Wortham's method was used because of its adaptability with this problem and it is shown that the solutions achieved is much faster than those resulted by Adlakha and Kulkarni's method.

Harrison [20] in a primarily theoretical study, introduced a queueing theoretical model of an assembly operation. Clearly, the arrival process to the manufacturing stations prior to an assembly station is Poisson with the rate of \( \lambda \). Som et al. [22] showed that the input stream the kitting operation is a Markov renewal process. They in a stochastic assembly system have shown that arrival stream of kits (a set of parts which are all needed to perform the assembly) into assembly station, follows Poisson distribution with \( \lambda \) as rate, as long as the parts arrive as a Poisson process of rate \( \lambda \). when unlimited inter-station buffers in size exist, a good estimation of rate \( \lambda \) is gained. In accordance with Song et al. [11], in multistage systems, development of an exact and direct analytic method for allocating delivery due date, because of difficulties in obtaining completion time distribution, is impossible. Therefore the completion time distribution function used in this paper is an approximate one.

### 2.1. NETWORK-BASED DUE DATE SETTING

#### 2.1.1. ASSUMPTIONS

The assembly system is modeled as an open queueing network, where each service station settled in a node of the network represents a manufacturing or assembly operation. One type of product is produced by the system (single product). Processing times of manufacturing and assembly operations are exponentially distributed (including set up times on the service station). Each product consists of a number of different raw parts, which should be processed and assembled to each other. The product order arrives at the source node according to a Poisson process. Then, it is split into some raw materials and each raw part goes to the particular manufacturing station for its first manufacturing operation, instantaneously, according to the same Poisson process with the rate \( \lambda \). If there are parts for being processed, it queues up. After the completion of processing at a manufacturing station, it goes to another manufacturing station to be processed in its routing sequence of manufacturing operations. After completing the manufacturing operations of each part, it is assembled to some other parts in an assembly station. The finished product leaves the system at sink node of the queueing network. Each service station with only one incoming arc represents a manufacturing station. Each service station with more than one incoming arcs represents an assembly station. Having completed one assembly, the server immediately begins another if at least one input item of each part is available. Otherwise, one or more parts have to wait for the last one to arrive (synchronization loss). The reader will recognize that each assembly station is a multi-input generalization of M/M/1 queue. Its salient feature is a very special type of batch servicing, each batch containing exactly one customer of each part. In each service station, there is either one or infinite number of machines with exponentially distributed processing times. Each part has characteristics, which is statistically independent of other parts. The processing time at each service station is independent of preceding processing times. In a non-preemptive setting, once the processing of a job starts, it must be completed without any interruption. Therefore, there are no interruptions due to breakdowns, maintenance, or other such cases.
Customers can be divided into different classes based on the revenues (or margins) they generate, or based on their demand characteristics. In this paper, one kind of customer is studied. Service discipline is FCFS. All inter-station buffers are infinite. The time needed to cross the inter-section buffers may be much greater than the service time and there is no reason to claim that its effect on the system will be negligible. It has been considered deterministic by some papers related to due date (see [16]). The queueing network is in the steady-state. The service rates are the known parameters. The service rate of each service station with one machine is greater than that of product order.

2.1.2. Completion Time Distribution Function

This paper defines the due date of order as the flowtime plus its arrival time; both of them are assumed to be stochastic variables. By applying longest path in queueing networks, the flowtime distribution function is obtained then with convolution, the completion time distribution function is computed.

In network-based due date method, each queueing network is transformed into an equivalent stochastic network with independent and exponentially distributed arc lengths. Then, the distribution function of longest path is computed from the source node to the sink node of this stochastic network by applying the method developed by Hartley and Wortham [27]. The main steps of proposed method are as follows:

Step 1 Compute the density function of the time spent in each service station.

Step 1.1 If there is one machine in the ith service station, then the distribution of waiting time (processing time plus waiting time in queue) in this M/M/1 queueing system is

\[ w_i(t) = (\mu_i - \lambda)e^{-(\mu_i - \lambda)t}, \quad t > 0 \]  

where \( \lambda \) and \( \mu_i \) are the arrival rate and the service rate of this queueing system, respectively. Therefore, the distribution of waiting time in the ith service station would be exponential with parameter \((\mu_i - \lambda)\) [21].

Step 1.2 If there are infinite machines in the ith service station, then the waiting time in this M/M/\( \infty \) queueing system will be exponentially distributed with parameter \( \mu_i \), because there is no queue [21].

Step 2 Transform the queueing network into an equivalent stochastic network by replacing each node that contains a service station with a stochastic arc whose length is equal to the time spent in the particular service station [23].

In this step, it is explained how to replace node \( k \) in the network of queues, which contains a queueing system, with a stochastic arc. Assume that \( b_1, b_2, \ldots, b_n \) are the incoming arcs to this node and \( d_1 \) is the outgoing arc from it. Then, this node is substituted by arc \((k', k'')\), whose length is equal to the system waiting time for the particular queueing system. Furthermore, all arcs \( b_i \) for \( i=1, \ldots, n \) end up with \( k' \) while the arc \( d_1 \) starts from node \( k'' \). Therefore, the network of queues is transformed into a stochastic network. The indicated process is opposite of the absorption an edge \( e \) in a graph \( G \) in graph theory \((G.e)\), see [23] for more details. After transforming all such nodes to the proper stochastic arcs, the queueing network is transformed into an equivalent stochastic network.

Step 3 Transform the original stochastic network, obtained in Step 2, into a new one with exponentially distributed arc lengths.

Step 4 Compute the distribution function of longest path in the new stochastic network obtained in Step 3, following the method of Hartley and Wortham [27].

After computing flowtime function in this way, according to \( d = r + f \) distribution function of product completion time can be achieved by convolution formula [24].

\[ F_D(d) = \int G_R(d-f) dG_F(f) = \int G_R(d-f) dG_F(f) \]

So distribution function of product completion time is as follows:

\[ F_D(d) = \int_0^d G_R(d-f) G_F(f) df \]

where \( D \), \( F \) and \( R \) are delivery due date random variable, flowtime random variable and arrival time random variable, respectively. \( G_F \) and \( G_R \) are cumulative distribution function of flow time and arrival time, respectively.

2.2. Due Date Setting for Assembly Systems: Meeting Service Target

Due date setting is often gained through a trade-off between earliness and tardiness costs. Earliness penalty is holding costs and other costs associated with completing a job before its due date. Tardiness penalty may be invoked if the product is completed after the due date agreed with the customer. This penalty includes loss of goodwill,
expedited shipment, etc. In many situations, a job completion time that deviates further and further from the due date should be assessed a much higher penalty and not just be charged a penalty that is linearly proportional to the deviation. For instance, contracts of aerospace industry may impose tardiness penalties as high as one million dollars per day on subcontractors for aircraft components [25]. The quadratic tardiness penalty is useful in cases where the tardiness cost function is assumed quadratic, although Song et al. considered linear tardiness penalty. For the asymmetric earliness and tardiness (AET), the expected cost is:

$$C(d') = \int_0^\infty e^{-c} (d' - d) F_d'(d) \, dd + \int_0^\infty e^{-c} (d - d') F_d'(d') \, dd$$

(3)

The other important way of assigning due dates is by specifying the probability of completion before due date that is the service target. Due dates are set to match specified service targets. Due date is assigned so that the probability of completing the product before the due date is equal to the service target. A due date \(d\) can be calculated to achieve a specified service target \(p\) as the solution of:

$$F_d(d) = p$$

(4)

For a multi stage assembly system, the distribution is expressed analytically.

In this paper, with quadratic tardiness cost after differentiating and simplifying Eq. 3, the following equation will be gained:

$$F_d(d') = \frac{2e^{-c} \left( \int_d^{\infty} dF_d'(d') \, dd - d' \right)}{c - 2c - d'}$$

(5)

The service targets set to produce plans depends on specific context and assessment of risks. Service targets may be set to meet cost criteria. The above relation indicates that the service target gained from that equation presents a due date plan with the same cost of the procedure explained earlier; minimization of earliness and tardiness costs. If the aim is to minimize the total of earliness and tardiness costs, then Song et al. [11] show that with \(c^+\) and \(c^-\) as the earliness and tardiness penalty costs per unit time, the minimum total cost occurs when the service target \(p = c^-/c^- + c^+\) is used in generating the due date plan. If the cost function is considered to be linear, the minimum total cost occurs when the service target \(p = c^-/c^- + c^+\) is applied which is in full agreement with the work of Song et al. [11].

3. NUMERICAL EXAMPLE

Consider the dynamic assembly system depicted in Figure 2. Service rates and arrival rate are (cost unit is in dollar and time unit is in day):

\(\lambda = 5, \mu_1 = 7, \mu_2 = 6, \mu_3 = 8, \mu_4 = 6, \mu_5 = 9\)

All service stations are \(M/1/M\). The transport times between the service stations settled in nodes 2 and 4, and also between those settled in nodes 3 and 5 are independent exponentially distributed random variables with the parameters \(\lambda(2,4) = 3\) and \(\lambda(3,5) = 1\).

The service station settled in node 5 is assembly station. After assembling, finished product is in hand. The transport times between the other service stations are zero. Now, the queueing network is transformed into the equivalent stochastic network with independent and exponentially distributed arc lengths which is shown in Figure 3.

In this network, the arcs 1, 2, 3, 6 and 7 indicate the waiting times in the \(M/1/M\) queueing systems settled in nodes 1, 2, 3, 4 and 5 of the queueing network with the parameters 2, 1, 3, 1, 4. The arcs 4 and 5 indicate the transport times between the service stations settled in nodes 2 and 4 with parameter 3, and also between the service stations settled in nodes 3 and 5 of the queueing network with parameter 1.

According to the proposed algorithm, the flow time distribution in the multistage assembly is computed as follows:
The arrival time is assumed to be of uniform distribution between zero and 10. The completion time distribution function, according to (6), is then computed as follows:

\[ F_{D}(d) = \int_{0}^{d} G_{F}(d - f) \frac{1}{10} (u(f) - u(f - 10)) df \]

in which \( G_{F}(f) \) is the flow time distribution of the product, \( u(\bullet) \) is the stepwise function and \( F_{D}(d) \) is the completion time distribution function.

\[
F_{D}(d) = \begin{cases} 
\frac{1}{10} \int_{0}^{d} G_{F}(d - f) df, & 0 < d \leq 10 \\
\frac{1}{10} \int_{0}^{10} G_{F}(d - f) df, & d > 10 \\
0, & d \leq 0 
\end{cases}
\]

Now in this instance the completion time distribution function could be obtained in (9). The cumulative graph of completion time is depicted in Figure 4. Table 1 compares the due dates set for different specified probabilities of completion.

\[
F_{D}(d) = \begin{cases} 
0.305 + 0.1d - 0.06e^{-5d} + 0.1823e^{-4d} - 0.525e^{-2d} + 0.633e^{-d} + 0.6e^{-3d} \\
-0.075d^2e^{-4d} - 0.187de^{-4d} + 0.2de^{-2d} - 0.45de^{-2d} + 0.0083e^{-6d} & 0 < d < 10 \\
1 - 0.06e^{-5d} + 0.1823e^{-4d} - 0.525e^{-2d} + 0.633e^{-d} + 0.6e^{-3d} + 0.0667e^{-3d} \\
-0.075d^2e^{-4d} - 0.1875de^{-4d} + 0.2de^{-2d} - 0.45de^{-2d} + 0.0083e^{-6d} - 0.6de^{-3d + 30} & d \geq 10 \\
+0.075d^2e^{-4d + 40} - 1.3125de^{-4d + 40} - 0.2de^{-d + 10} + 0.45de^{-2d + 20} + 0.06e^{-5d + 50} \\
+5.442e^{-4d + 40} - 3.975e^{-2d + 20} + 1.3667e^{-d + 10} + 5.933e^{-3d + 30} - 0.0083e^{-6d + 60} & d \leq 0 
\end{cases}
\]

Figure 4. Completion time distribution function

<table>
<thead>
<tr>
<th>Service target</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due date</td>
<td>8.049</td>
<td>9.053</td>
<td>10.055</td>
<td>11.06</td>
<td>12.23</td>
<td>13.16</td>
</tr>
</tbody>
</table>

Figure 5 indicates good agreement between the results of the approximation method of system flow time and simulation results. The differences are of the magnitude of 3.38%.
4. COMPUTATIONAL RESULTS

Hartley and Wortham's algorithm, as mentioned before, calculates the distribution function of longest path through reducing the network by serial and parallel branches and includes rather smooth and clear calculations. However, Kulkarni and Adlakha's algorithm calculates the function through the concept of continuous-time Markov chain and solving differential equation derived from Chapman-Kolmogorov’s equations. In the latter, all the elements in the upper triangle matrix Q should be computed using the logic behind it. This process last several hours for an average scale network. For three test problems each contained three parameter sets, Hartley and Wortham's algorithm is compared with Kulkarni and Adlakha's method. A computer program was written to put the algorithm into action. The program was written in MATLAB 7.1 environment and run on a computer having an AMD Athlon 64 3500+ processor (2.21 GHz) and 1.00 GB of RAM and under windows XP professional. In the worst case (instance 3) it took 104.73 seconds to achieve the solution while under equal situation, Hartley and Wortham’s algorithm took only 1 second (Table 2). This shows the superiority of Hartley and Wortham’s algorithm in both time and quality.

Table 2. Running time (second) for the test problems

<table>
<thead>
<tr>
<th>Parameter set</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kulkarni's method</td>
<td>1.11</td>
<td>4.01</td>
<td>77.16</td>
</tr>
<tr>
<td>Hartley and Wortham's method</td>
<td>0.40</td>
<td>0.56</td>
<td>0.55</td>
</tr>
</tbody>
</table>

For being sure about the accuracy of the simulation model, three test problems each with three different sets of parameters were carried out for 1000 runs. The developed method seems to work pretty well under our assumptions. The max value for difference between flowtime distributions in corresponding analytical and simulation procedures is 6.11 %. Results are summarized in Table 3 containing the standard errors for estimated points.

Table 3. Difference and standard error of validation for the test problems

<table>
<thead>
<tr>
<th>Parameter set</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum difference</td>
<td>3.38</td>
<td>5.2</td>
<td>3.29</td>
</tr>
<tr>
<td>Standard error</td>
<td>1.05</td>
<td>1.6</td>
<td>1.76</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS AND RECOMMENDATIONS

A product due date assignment procedure has been developed for multistage products with stages of manufacture and assembly. First, by applying the longest path analysis in queueing networks, the manufacturing lead time distribution is computed. Then, the product completion time distribution function is computed. The optimal due date was calculated by meeting service target. Kulkarni and Adlakha’s algorithm has been used in different applications [19]. It was shown that the solutions achieved by Hartley and Wortham is significantly much faster than those obtained by Adlakha and Kulkarni's method, so this method can be used for real life application in production systems.
The method could be further improved to take into consideration bargaining and customer opinions. Another area for the future research is applying the multi objective techniques such as goal programming. This area and several other related areas are being currently studied by the authors.

REFERENCES

Hybrid metaheuristics for a SDST parallel machines scheduling problem with bi-objectives: makespan and sum of the earliness and tardiness

J. Behnamian¹, S.M.T. Fatemi Ghomi¹, a, M. Zandieh b

¹Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, Iran
bDepartment of Industrial Management, Management and Accounting Faculty, Shahid Beheshti University, Tehran, Iran

ABSTRACT

In the literature of multi-objective problem there are different algorithms for solving different optimization problems. This paper presents a min-max multi-objective procedure for a dual-objective, namely makespan, and sum of the earliness and tardiness of jobs in due window problems, simultaneously. In formulation of min-max method when this method is combined with the weighting method, the decision-maker can have the flexibility of mixed use of weights and distance parameter in yielding a set of Pareto-efficient solutions. This research extends the new hybrid metaheuristic (HMH) for solving parallel machines scheduling problems with sequence-dependent setup times which comprises three components: an initial population generation method based on an ant colony optimization (ACO), a Simulated annealing (SA) as an evolutionary algorithm employs certain probability to avoid becoming trapped in a local optimum, and a variable neighborhood search (VNS) which involves three local search procedures to improve the population. In addition, two VNS-based hybrid metaheuristics, which are a combination of two methods, SA/VNS and ACO/VNS, are also proposed for solving the addressed scheduling problems. The non-dominated sets obtained from each of algorithms are compared in terms of various indices, and the computational results show that the proposed algorithm is capable of producing a number of high-quality Pareto optimal scheduling plans. Aside, an extensive computational experience is carried out in order to analyze the different parameters of the algorithm.

1. INTRODUCTION

The problem of scheduling jobs in parallel machines has been extensively investigated during the last decades. Most of the contributions consider a single optimization criterion, although in practice the decision maker (DM) often faces several (usually conflicting) criteria. Therefore, multi-criteria scheduling is considered to be a key research area in order to address the gap between scheduling research and practice [1]. For reflecting real-world situations adequately, this paper formulates multi-objective scheduling problems as two-objective ones which simultaneously minimize the maximum completion time (i.e., makespan), and the minimization of sum of the earliness and tardiness (ET) of jobs in due window problems. No penalty is incurred if a job is completed at any time within its due window.

We consider a bi-objective scheduling of n jobs on an identical parallel machines with sequence-dependent setup times (SDST). Much of the research on operations scheduling problems has ignored setup times but this assumption is commonly observed in various industrial settings including printing, textile, pharmaceutical, chemical and metallurgical industries. Das et al. [2] and Gravel et al. [3] document are examples of sequence-dependent setups in the various industries.

Research in the area of scheduling parallel machines with multiple objectives is limited. Suresh and Chaudhuri [4] proposed a tabu search method to solve a bi-criteria problem involving minimizing maximum completion time and maximum tardiness for unrelated parallel machines. Lin and Liao [5] consider the identical parallel machines problem with makespan minimization subject to minimum total flow time. Angel et al. [6] studied parallel machines scheduling problem with the objectives of minimizing the makespan and total weighted flowtime of jobs. They show that it is possible to construct in polynomial time an approximate Pareto curve whenever the number of machines is a constant. Cochran et al. [7] propose a two-stage multi-population genetic algorithm to solve the parallel machines scheduling problem with two objectives: makespan and total weighted tardiness. They use the genetic algorithm to assign jobs to machines and then sequence them based on the different heuristics proposed in that paper for each objective. Moreover, Yu et al. [8] propose a two stage

¹ Corresponding author: E-mail addresses: fatemi@aut.ac.ir
Lagrangian Relaxation Heuristic method to solve the unrelated parallel machines scheduling with consideration of the three objectives: total completion time, total tardiness, and total number of tardy jobs.

The parallel machines scheduling problem of minimizing total earliness and tardiness with a common weight for earliness and tardiness, respectively, as well as given sequence-dependent setup times for the jobs is treated by Sivrikaya-Serifoglu and Ulusoy [9]. Machines are either identical or uniform. For this problem, two genetic algorithms were given and tested on problems with up to 60 jobs and four machines. Heady and Zhu [10] minimized the total weighted deviation of the completion times from the given due dates, where machines may be unrelated and sequence dependent setup times are given. Lam and Xing [11] gave a short review of new developments of parallel machines scheduling problems associated with setup times of JIT production. Preemption with setup and capacitated machine scheduling. Balakrishnan et al. [12] used a variable neighborhood search formulation that had substantially fewer zero one variables than typical formulations for scheduling problems of this type. Kim et al. [13] attempted to solve the unrelated parallel machines and sequence-dependent setup times with new approach in 2002. In this particular case, the setup times are not machine dependent and each job is composed of \( n \) items. In the line of batch scheduling problems, Kima et al. [14] could be cited, where several search heuristics are presented and tested considering parallel machines. Logendrana et al. [15] deal with the minimization of the makespan, the problem is known to be a NP-hard problem [17].

Thus, the more complex case of bi-objective scheduling problem is also NP-Hard, and consequently, devising a (meta)heuristic to solve this problem is highly desirable. The pur pose of this work is to evaluate the effectiveness of integrating such a specific subset of features into a configurable hybrid metaheuristic (HMH), at least for the class of difficult scheduling problems here considered, consisting of the generalization of the SDST parallel machines scheduling problem. The remainder of this paper is organized as follows. Section 2 is problem description. Section 3 introduces the proposed hybrid algorithm and multi-objective technique that is used in this paper. Section 4 presents the two VNS-based hybrid algorithms used in the experiments. Section 5 presents and compares the hybrid algorithm with version of VNS-based algorithms and VNS algorithm presented recently for SDST parallel machines scheduling. Finally, Section 6 states our conclusions and further researches on this topic.

2. PROBLEM DESCRIPTION

This paper studies the problem of scheduling a set of \( n \) independent jobs on \( m \) identical parallel machines. When a job \( k \) is processed after job \( j \) a setup time \( s_{kj} \) is incurred. All the jobs have a due window \([d_{ij}, d_{ij}]\), where \( d_{ij} \) is the earliest due date and \( d_{ij} \leq d_{ij} \) the latest due date. Furthermore, this paper assumes that the due window is given in advance, i.e., both \( d_{ij} \) and \( d_{ij} \) are prespecified. Each job \( j \in n \) is ready at time \( 0 \) and has a known processing time \( p_j \). If a job \( j \in n \) is completed before \( d_{ij} \), it will incur an earliness penalty and if a job \( j \in n \) is completed after \( d_{ij} \), it will incur tardiness penalty.

3. METHODOLOGIES

3.1. MULTI-OBJECTIVE METHOD: NO PREFERENCE ARTICULATION

Multi-objective optimization problems usually solved by scalarization [18]. Scalarization means converting the problem with multiple objectives into a single objective optimization problem or a family of single objective optimization problems. When a multi-objective optimization problem has been scalarized, methods developed for single objective optimization can be used to solving the problem. The objective function of the single objective problem is called a scalarizing function [19].

The following achievement scalarizing function is minimized subject to the constraint of the problem:

\[
\min \left[ \sum_{i=1}^{n} \left( \frac{f_i(x) - f_i^*}{f_i^*} \right)^p \right]
\]

s.t. \( X \in S \)

\( 1 \leq p \leq \infty \)

The exponent \( p \) gives different ways of calculating the distance. The most frequently used values for \( p \) are 1 for the simplest formulation, 2 for the Euclidean distance, and \( \infty \) for Tchebycheff norm. The difficulty with this approach is finding the value for \( p \) that will maximize the satisfaction of the DM. In this paper, we as sume values for \( p \) are 1, 2, and 100 (as large number instead of \( \infty \)).
In this method the output is just one point on the Pareto front, and the DM has to accept it as the final solution. The optimum, in the min-max sense, gives a solution that treats all the objectives on terms of equal importance, and presents the advantage of being very efficient and easy to implement [20]. Furthermore, when the min-max approach is combined with the weighting method, and by changing the exponent in the distance formulation, the set of Pareto (non-dominating) solutions could be generated for both convex and non-convex problems. This technique was tested with multi-objective engineering design problems found in the literature [21]. For this reason, min-max method with the weighting method was combined to generate various Pareto solutions. Thus, the objective function value is defined by:

$$
\min \left[ w \left( \frac{f_1(x) - f_1^*}{f_1^*} \right) + (1 - w) \left( \frac{f_2(x) - f_2^*}{f_2^*} \right) \right]^{\frac{1}{p}},
$$

where $0 \leq w \leq 1$. $W$ denotes the weight (or relative importance) given to $C_{\max}$ and ET penalties in due window problems. Note that, all instances are solved using 10 different seeds for each algorithm and the minimum solution in all runs is used as $f^*$ for each objective. Then they are replaced in equations (1, 2).

3.2. THE HYBRID METAHEURISTIC APPROACH

The steps of the HMH algorithm are as follows:

Step 1: Generating the initial solution with an ant-sequence: The method starts with generating initial solutions by employing ACO.

Step 2: Intensification phase using VNS. The concept borrowed by the HMH from the VNS is that varying the neighborhood structure during the search process could facilitate the avoidance of traps and enlarge the search scope. In this phase first, the ET value for each neighborhood solution is calculated. VNS works by performing movements that upgrade the solutions. So, the solution that has the minimal ET value is selected as a move. The three neighborhoods employed in our algorithm are defined below:

1. Job swaps on one machine: one machine is chosen and all possible job swaps are considered.
2. Job swaps between two different machines: two machines are chosen and all possible job swaps from different machines are considered.
3. Job transfers from one machine to another: one machine is chosen and all possible job movements from this machine to any other are considered.

Step 3: diversification phase via a SA and VNS. In HMH a SA only is applied to diversification the solution. The asymptotic convergence of SA to a global optimum has been proved, even if such a result has only a theoretical relevance [22]. At iteration, a solution $p_0$ is selected from the neighborhood based on the VNS technique. When each movement is available, unlike VNS techniques, SA executes a second phase where moving from the current solution to a worse solution is allowed with the probability, with the expectation that this movement will eventually guide to a better solution. However, if $p_0$ is not accepted, then neighborhood structure is changed in the same way as in VNS technique (shaking function).

Step 4: Global pheromone update phase and ant generation for next solution by means of ACO. This phase is performed after each ant has completed its schedule. In order to make the search more directed, the global pheromone update with intended to provide a larger amount of pheromone to better schedules.

The basic proposed hybrid algorithm structure is designed as shown in following procedure.

**Hybrid metaheuristic procedure**

1. $R_{\text{time}} \leftarrow$ Set run time()
2. while run time $< R_{\text{time}}$ do
3. $S^* \leftarrow$ Generate initial solution()
4. $l \leftarrow 1$
5. for each ant do
6. $S \leftarrow S^*$;
7. $k \leftarrow 0$
8. $T_i \leftarrow$ Set initial temperature()
9. $T_f \leftarrow$ Set final temperature()
10. while current temperature $> T_f$ do
11. Shake procedure: find a random solution $S' \in N_i (S)$;
12. Perform a local search on $N_i (S')$ to find a solution $S''$;
13. if $f(S') \leq f(S^*)$ then $S^* \leftarrow S''$;
14. $l \leftarrow 1$
15. else
16. $l \leftarrow l + 1$
17. $T_i \leftarrow T_i * \alpha$
18. $T_f \leftarrow T_f / \beta$
19. break
20. $T_i \leftarrow T_i / \gamma$
21. $T_f \leftarrow T_f * \delta$
22. while $T_i > T_f$ do
23. $T_i \leftarrow T_i / \eta$
24. $T_f \leftarrow T_f * \zeta$
25. end while
26. $S \leftarrow S' * (1 - \rho)$
27. $S \leftarrow S' + \sigma$
28. end while
29. $S \leftarrow S^*$
30. end for
31. end while
32. end while
33. end while
34. end for
4. VNS-BASED HYBRID ALGORITHMS

This section illustrates and discusses in detail the proposed two VNS-based algorithms with reference to the following steps that summarize the HMH procedure. In each algorithm the initial solution is generated randomly then combine one the main metaheuristic ACO and SA algorithm with variable neighborhood search. Basically, these hybrids metaheuristic consist of two parts: the construction of a main algorithm and a local search that make with VNS algorithm. Both the phases are repeated in every iteration.

5. EXPERIMENTAL DESIGN

In this section the proposed hybrid algorithms with the VNS algorithm which proposed by Rocha et al. [16] for the SDST parallel machines scheduling with makespan objective are compared. In this work VNS algorithm was adapted to the multi-objective problem. The proposed algorithms and VNS algorithm are coded in C++ and run with an Intel Pentium IV dual core 2.5 GHz PC at 896 MB RAM under a Microsoft Windows XP environment.

5.1. DATA GENERATION AND SETTINGS

The problem data can be characterized by three factors, and each of these factors can have at least two levels. These levels are shown in Table 1.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of jobs</td>
<td>6, 30, 100</td>
</tr>
<tr>
<td>Machine distribution</td>
<td>Constant: 1, Variable: Uniform(1, 4), Uniform(1, 10)</td>
</tr>
<tr>
<td>Processing times</td>
<td>Uniform(50, 70), Uniform(20, 100)</td>
</tr>
</tbody>
</table>

The setup times are uniformly distributed from 12 to 24 which are 20% to 40% of their processing time. The setup time matrices are asymmetric and satisfy the triangle inequality. The setup time characteristics follow Rios-Mercado and Bard [23].

5.2. STOPPING RULE

The stopping criterion used when testing all instances with the algorithms is set to a computational time (CPU time) limit fixed to \((m/2) \times (n+1)\) seconds. This stopping criterion not only responsive to number of parallel machines, but also is sensitive toward rise in number of jobs.

5.3. EVALUATION METRIC

After computation of objective value of each algorithm for its instances, the best solution obtained for each instance (which is noted \(Min_{sol}\)) by any of the four algorithms is calculated. Relative percentage deviation (RPD) is obtained by the following formula:

\[
RPD = \frac{Alg_{sol} - Min_{sol}}{Min_{sol}}
\]  

(3)

where \(Alg_{sol}\) is the objective value obtained for a given algorithm and instance.

5.4. EXPERIMENTAL RESULTS

For all problem sizes, instances was tested with \(w \in \{0, 0.2, 0.5, 0.8, \text{ and } 1\}\) in the objective function. Table
2 shows the results of the experiments for two subsets, averaged for each one of the \( n, m \) and \( w \) configurations. The results demonstrate that there is a clear statistically significant difference between performances of the algorithms for \( w \in \{0, 0.2, 0.5, 0.8, \text{and} 1\} \) with \( p=1 \). Figure 1 shows the means plot and LSD intervals (at the 95% confidence level) for four algorithms. Note that, the results of the LSD method helped us identify robust HMH that performed significantly better than others.

![Figure 1: Plots of \( \text{RPD} \) for the type of algorithm factor and \( w \)](image)

5.5. **ANALYSIS OF CONTROLLED FACTORS**

**ANALYSIS OF PROBLEM SIZE FACTOR (NUMBER OF JOBS)** In order to see the effect of number of jobs on two algorithms, a two ways ANOVA is applied. Means plot and LSD intervals (at the 95% confidence level) for the interaction among the factors of type of algorithm, number of jobs, \( w \), and \( p \) are shown in Figure 2. As it could be seen, in the all cases of \( n = 6, n = 30 \) and \( n = 100 \) the HMH works well than all other algorithms.

**ANALYSIS OF M FACTOR (NUMBER OF MACHINES)** Another two ways ANOVA and LSD test are applied to see the effect of magnitude of machines, \( w \) and \( p \) on quality of the algorithms. Figure 3 shows the results. As it could be seen, in the total cases the HMH works better than others.

![Figure 2: Plots of \( \text{RPD} \) for the interaction among the type of algorithm, number of jobs and \( w \)](image)

![Figure 3: Plots of \( \text{RPD} \) for the interaction among the type of algorithm, magnitude of machines and \( w \)](image)
Table 2 Average relative percentage deviation (RPD) for algorithms grouped by n, m and w with p=1

<table>
<thead>
<tr>
<th>Problem size</th>
<th>Algorithm &amp; weight</th>
<th>t=0</th>
<th>t=0.2</th>
<th>t=0.5</th>
<th>t=0.8</th>
<th>w=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>n, m, w</td>
<td>HMH</td>
<td>SAV</td>
<td>ANV</td>
<td>VNS</td>
<td>HMH</td>
<td>SAV</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.74</td>
<td>0.57</td>
<td>0.62</td>
<td>0.19</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>0.09</td>
<td>0.75</td>
<td>0.46</td>
<td>0.51</td>
<td>0.11</td>
<td>0.96</td>
</tr>
<tr>
<td>U(1,4)</td>
<td>0.14</td>
<td>0.71</td>
<td>0.47</td>
<td>0.51</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>U(1,6)</td>
<td>0.11</td>
<td>0.72</td>
<td>0.48</td>
<td>0.53</td>
<td>0.15</td>
<td>0.78</td>
</tr>
<tr>
<td>6 Job</td>
<td>0.11</td>
<td>0.73</td>
<td>0.51</td>
<td>0.56</td>
<td>0.16</td>
<td>0.84</td>
</tr>
<tr>
<td>1</td>
<td>0.09</td>
<td>0.56</td>
<td>0.77</td>
<td>0.69</td>
<td>0.39</td>
<td>1.33</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.55</td>
<td>0.78</td>
<td>0.77</td>
<td>0.34</td>
<td>0.71</td>
</tr>
<tr>
<td>30</td>
<td>0.05</td>
<td>0.78</td>
<td>0.74</td>
<td>0.75</td>
<td>0.33</td>
<td>0.99</td>
</tr>
<tr>
<td>U(1,4)</td>
<td>0.16</td>
<td>0.53</td>
<td>0.78</td>
<td>0.75</td>
<td>0.27</td>
<td>1.03</td>
</tr>
<tr>
<td>U(1,10)</td>
<td>0.12</td>
<td>0.60</td>
<td>0.80</td>
<td>0.74</td>
<td>0.22</td>
<td>1.44</td>
</tr>
<tr>
<td>30 Job</td>
<td>0.10</td>
<td>0.62</td>
<td>0.78</td>
<td>0.73</td>
<td>0.29</td>
<td>1.34</td>
</tr>
<tr>
<td>1</td>
<td>0.07</td>
<td>0.73</td>
<td>0.63</td>
<td>0.59</td>
<td>0.22</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>0.79</td>
<td>0.68</td>
<td>0.56</td>
<td>0.16</td>
<td>0.58</td>
</tr>
<tr>
<td>100</td>
<td>0.15</td>
<td>0.69</td>
<td>0.76</td>
<td>0.56</td>
<td>0.01</td>
<td>0.55</td>
</tr>
<tr>
<td>U(1,4)</td>
<td>0.06</td>
<td>0.77</td>
<td>0.72</td>
<td>0.60</td>
<td>0.13</td>
<td>0.62</td>
</tr>
<tr>
<td>U(1,10)</td>
<td>0.06</td>
<td>0.80</td>
<td>0.68</td>
<td>0.55</td>
<td>0.13</td>
<td>0.95</td>
</tr>
<tr>
<td>100 Job</td>
<td>0.08</td>
<td>0.76</td>
<td>0.68</td>
<td>0.56</td>
<td>0.13</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Average: 0.10 0.70 0.66 0.62 0.19 1.00 1.39 1.48 0.14 1.17 1.29 1.19 0.23 1.04 1.36 1.26 0.02 0.05 0.06 0.07
5.6. EVALUATION OF NON-DOMINATED SOLUTION

In this section a quantitative measure is needed for compared sets of Pareto solution. For this reason several approaches in multi-objective literature. This paper proposed four indices as follows:

1. **MID** (mean ideal distance): The closeness between Pareto solution and ideal point (0, 0),
2. **SNS**: The spread of non-dominance solution,
3. **RAS**: The rate of achievement to two objectives simultaneously, and
4. **ALC**: Area under linear regression curve that fits to Pareto solution data with Excel Trendline function ($S_1$ for hybrid algorithm and $S_2$ for VNS algorithm).

The equation of **MID** is defined as follows:

$$MID = \frac{\sum_{i=1}^{n} c_i}{n}$$

where $n$ is the number of non-dominated set and $c_i = \sqrt{f_1^2 + f_2^2}$. The lower value of **MID**, the better of solution quality we have.

**SNS**, as diversity measure, can be expressed by the following equation:

$$SNS = \sqrt{\frac{\sum_{i=1}^{n} \left(MID - c_i\right)^2}{n-1}}$$

The higher value of **SNS**, the better of solution quality we have (more diversity in obtained solution). Another evaluation metric which was proposed in this paper is **RAS**. The equation of **RAS** is represented as equation (6).

$$RAS = \frac{\sum_{i=1}^{n} \left(\frac{f_1}{F_i} - 1\right) + \left(\frac{f_2}{F_i} - 1\right)}{n}$$

where $F_i = \min\{f_{1i}, f_{2i}\}$. The lower value of **RAS**, the better of solution quality we have.

<table>
<thead>
<tr>
<th>Problem size</th>
<th>Algorithm</th>
<th>MID</th>
<th>SNS</th>
<th>RAS</th>
<th>ALC</th>
</tr>
</thead>
<tbody>
<tr>
<td>30×8</td>
<td>HMH</td>
<td>0.89244</td>
<td>0.95640</td>
<td>0.23101</td>
<td>0.19366</td>
</tr>
<tr>
<td>100×8</td>
<td>HMH</td>
<td>0.64870</td>
<td>0.76163</td>
<td>0.01245</td>
<td>0.04781</td>
</tr>
</tbody>
</table>

5.7. NON-DOMINATED SOLUTIONS

Figure 4 presents the non-dominated solutions of a single run by multi-objective HMH and multi-objective VNS (for 2 sizes (30×8), and (100×10)). The results show that the proposed algorithm works effectively in both size problems.

6. SUMMARY

In this research to deal with bi-objective optimization the min-max technique is used to solve parallel machine scheduling problems in which they are desired to simultaneously minimize two criteria, maximum completion time and sum of the earliness and tardiness of jobs in due window problems. The core of our method is the novel idea of combining the min-max scalarizing function with the weighting method that he laps to
generate many Pareto and diverse solutions as possible. Several hybrid metaheuristics were proposed for solving addressed problem. The main hybrid metaheuristic that is called HMH has three unique features: It population-based evolutionary searching ability belonging to ACO, it ability to balancing exploration and exploitation belonging to SA, and it local improvement ability belonging to VNS. Experiments with three proposed hybrid metaheuristics and VNS algorithm proposed in a lately published demonstrated that one of them, namely HMM, typically outperforms all others, and it generates the most efficient set of solutions. The future work is to change the objectives that are used in this paper and develop effective hybrid algorithm for other types of scheduling problems.

REFERENCES

Early/tardy scheduling with sequence-dependent setup times on hybrid flowshop using a novel hybrid metaheuristic

J. Behnamian a, S.M.T. Fatemi Ghomi 1, a, M. Zandieh b

a Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, Iran
b Department of Industrial Management, Management and Accounting Faculty, Shahid Beheshti University, Tehran, Iran

ABSTRACT

This paper presents an efficient hybrid metaheuristics for scheduling jobs in a hybrid flowshop with sequence-dependent setup times. The problem is to determine a schedule that minimizes sum of the earliness and tardiness of jobs. Since this problem class is NP-hard in the strong sense, there seems to be no escape from appealing to metaheuristic procedures to achieve near-optimal solutions for real life problems. This paper proposes the hybrid metaheuristic algorithm which comprises three components: an initial population generation method based on an ant colony optimization, a simulated annealing as an evolutionary algorithm employs certain probability to avoid becoming trapped in a local optimum, and a variable neighborhood search which involves three local search procedures to improve the population. A design of experiments approach is employed to calibrate the parameters of the algorithm. Results of computational tests in solving on 252 problems up to 100 jobs have shown that the proposed algorithm is computationally more effective in yielding solutions of better quality than the adapted random key genetic algorithm and immune algorithm presented previously.

1. INTRODUCTION

In a just-in-time environment, each job should be completed as close as possible to its due date. Traditionally, tardiness penalties due to delivery after a contractually arranged due date are considered. Missing a job’s due date may result in the loss of the customer or the need to compensate for the delay by a long time production or assemby line. On the other hand, producing a job much earlier than its due date may cause unwanted inventory and/or deterioration of the product. In such an environment, the length a job is tardy or early is important [1]. Therefore, tardiness together with earliness penalties are considered to meet the requests of practice and any schedule of a set of jobs should strive to minimize the total earliness and tardiness where tardiness reflects customer satisfaction and earliness measures inventory performance [2].

A hybrid flowshop model, commonly known as flexible flow line, allows us to represent most of the production systems. The process industries such as chemical, pharmaceutical, oil, food, tobacco, textile, paper, and metallurgical industries can be modeled as hybrid flowshops. A hybrid flowshop consists of a series of production stages, each of which has several facilities in parallel [3]. The flow of products in the plant is unidirectional. Hybrid flowshop models differ in the type and number of machines at the stages.

Scheduling problems with sequence-dependent setup times (SDST) are among the most difficult classes of scheduling problems. Sequence-dependent setup times of a hybrid flowshop system is even more challenging. Although there has been some reported progress, the understanding of sequence-dependent setups, however, is still believed to be far from being complete [4]. Therefore in this paper, sequence-dependent setup time restrictions are taken into account as well.

This paper proposes a hybrid algorithm to SDST hybrid flowshop problems. The paper has the following structure. Section 2 gives the brief literature review. Section 3 is the problem description. Section 4 introduces the proposed hybrid algorithm. Section 5 presents the experimental design which compares the results achieved by proposed hybrid algorithm with those achieved by past genetic algorithm and immune algorithm. Finally, section 6 is devoted to conclusions and future works.

2. LITERATURE REVIEW

Most of the studies deal with problems with throughput related measures, such as minimizing makespan or mean flow time, but there are not many studies on problems with performance measures related to due dates.

1 Corresponding author: E-mail addresses: fatemi@aut.ac.ir
Adenso-Diaz [5] proposed a hybrid algorithm based on Simulated Annealing (SA) and tabu search (TS) in which the solution of SA was used as initial one by TS method where the size of the neighborhood was limited, and they compared the solution obtained against a tabu search starting from the same initial solution without any restriction in the size of neighborhood. Brah [6] investigates performance of ten priority rules in terms of mean and maximum tardiness, and Lee et al. [7] suggest heuristic algorithms for hybrid flowshops with the objective of minimizing total tardiness. Allouati and Artiba [8] addressed the FFm/STsi problem with respect to the criteria of Cmax, Tmax, Tj, and Cj with machine unavailability intervals (due to breakdowns or eventive maintenance), where the transportation times between the stages are explicitly considered. Jungwattanakit et al. [9] considered the FFm/STsi/STsd problem for the case of unrelated parallel machines, where \( \theta \leq 1 \). They adapted the well known constructive heuristics and iterative heuristics (genetic algorithm and simulated annealing) for the pure flowshop problems.

While many solution procedures have been developed over many years to solve the scheduling problems, only some attempts have been made to develop procedures for the hybrid flowshop problem with sequence dependent setup times of jobs (see for instance the recent review on scheduling with setup times by Allahverdi et al. [10]). However, in most real world cases, the length of the setup depends on the job sequences, which is separable from processing. There seems to be published only three works addressing heuristics for flexible flowlines with sequence dependent setup times. Pugazhendhi et al. [11] addressed the FFm/STsi/\( \sum w_j f_j \) problem, where some jobs may have missing operations on some machines. They proposed a heuristic procedure to derive a non-permutation schedule from a given permutation sequence. Kurz and Askin [12] examined scheduling rules for SDST flexible flowlines. They explored three classes of heuristics.

Moreover, Kurz and Askin [13] formulated the SDST flexible flowlines as an integer programming model. Due to the difficulty of solving the IP model directly, they developed a random keys genetic algorithm. They created a lower bound to evaluate the heuristics. Zandieh et al. [14] proposed an immune algorithm, and showed that their algorithm outperforms the random keys genetic algorithm of Kurz and Askin [13].

3. DESCRIPTION OF PROBLEM AND NOTATION

Let \( g \) be the number of stages and \( m' \) be the number of machines in parallel at each stage \( t \). These machines are identical. There are \( n \) independent jobs that are available at time 0, to be scheduled without preemption. A machine can process at most one job at a time and a job can be processed by at most one machine at a time. Each product is processed at only one facility in each stage and at one or more stages before it exits the plant. The problem consists of assigning jobs to machines at each stage and sequencing the jobs assigned to the same machine so that some optimality criteria are minimized. The jobs can wait in between stages and the intermediate storage is unlimited. The setups are assumed to be sequence-dependent. This paper supposes that machines are initially setup for a nominal job 0 at every stage. Job \( n+1 \) exists at every stage only to indicate the end of the process, if needed. Let \( d_j \) be the due date for the job \( j \), the earliness and tardiness of job \( j \) are defined as \( E_j = \max(0, \text{Cj} - \text{dj}) \) and \( T_j = \max(0, \text{dj} - \text{Cj}) \), respectively, where \( \text{Cj} \) is the completion time of job \( j \). In particular, the objective function can be written as

\[
\text{Min} \quad Z = \sum_{j=1}^{n} (E_j + T_j) = \sum_{j=1}^{n} |d_j - \text{Cj}|. 
\]

When the objective function is written in this form, it is clear that earliness and tardiness are penalized at the same rate for all jobs.

4. THE HYBRID METAHEURISTIC APPROACH

The construction of hybrid metaheuristics is motivated by the need to achieve a good tradeoff between the capabilities of a heuristic to explore the search space and the possibility to exploit the experience accumulated during the search. When the HMH algorithm is proposed for hybrid flowshop problems, the first important step is the representation of a solution. Bean [15] introduced a random key approach for real-coded GA for solving sequencing problem. Subsequently, numerous researchers show that this concept is robust and can be applied for the solution of different kinds of scheduling problems. In this paper, the random key approach is used to represent a schedule incorporated with the hybrid methodology is applied to address problem. Let us no w dis scuss w hich a ccepts h ave been borrowed from Ant Colony Optimization (ACO), Variable neighborhood search (VNS) and SA.

4.1. ANT COLONY OPTIMIZATION

The ACO algorithm is implemented in this paper is basically the Ant Colony System (ACS), but it is pheromone values are limited to a bounded value. Explicit limits on the pheromone values prevent the probability for constructing a solution falls below a certain value greater than 0. This algorithm is called Max–Min Ant System (MMAS). The basic steps of the MMAS are presented below:
Step 1: Initialize the pheromone trails and parameters:
Step 2: While termination condition is not met do the following:
  • Construct a solution.
  • Improve the solution by local search.
  • Update the pheromone trail or trail intensity, denoted by $\tau_{ip}$, where $\tau_{min} \leq \tau_{ip} \leq \tau_{max}$.
Step 3: Return the best solution found.

In the context of application of the ACO algorithm to scheduling problems, $\tau_{ip}$ denotes the trail intensity (or desire) of setting job $i$ in position $p$ of a sequence. These trails form a kind of adaptive memory of previously found solutions, and are modified at the end of every iteration. It is to be noted that, for every job $i$, for any possible position $p$, a pheromone value is stored and updated in each iteration of the ACO algorithm. Hence, there are $n$ such values of $\tau_{ip}$.

4.1.1. Generating the Initial Solution with an Ant-Sequence

Initially, all $\tau_{ip}$’s are set equal to $\tau_{max}$, where $\tau_{max}$ is set equal to $1/((1 - \rho) \times Z_{best})$ and $\rho$ denotes the persistence of the trail. The term ‘$Z_{best}$’ refers to the best value of objective function that has been obtained so far. At first, $Z_{best}$ equals the objective-function value yielded by the seed sequence. Starting from a null sequence, ant colony algorithm makes use of trail intensities in order to determine the job to be appended in position $p$, where $1 \leq p \leq n$ and $n$ refers to the number of jobs to be scheduled. Although the trail intensity $\tau_{ip}$ changes with respect to every iteration in the ant colony algorithm, the iteration counter is omitted for the sake of simplicity of presentation. An ant starts constructing a sequence by choosing a job for the first position, followed by the choice of an unscheduled job for the second position, and so on. A dummy job ‘0’ is introduced on which an ant is set initially, and the construction of partial sequences begins, thereby leading to the build-up of a complete sequence by the ant, called an ant sequence. In the case of MMAS, the following procedure is used to choose an unscheduled job, say job $i$, probabilistically for position $p$.

Sample a uniform random number $u$ in the range (0,1).
If $u \leq (n-4)/n$ then
among the jobs that are not yet scheduled, choose the job with the maximum value of $\tau_{ip}$;
else
job $i$ is selected, from the set of first five unscheduled jobs as present in the best sequence obtained so far, for position $p$ by sampling from the following probability distribution:
$$E_{ip} = \frac{\tau_{ip}}{\sum_p \tau_{ip}}.$$  \hspace{1cm} (2)

where job $i'$ belongs to the set of first five unscheduled jobs, as present in the best sequence obtained so far. Note that when there are less than five jobs unscheduled, then all such unscheduled jobs are considered.

4.1.2. Updating the Trail Intensities

An ant-sequence is subjected to the position-based local search procedure to enhance the quality of solution. Let the objective-function value of this improved sequence be denoted by $Z_{current}$. Subsequently, trail intensities are updated as follows:
$$\tau_{new} = \begin{cases} \rho \times \tau_{old} + (1/ Z_{current}) & \text{if job } i \text{ is placed in position } p \text{ in the generated sequence;} \\ \rho \times \tau_{old} & \text{otherwise,} \end{cases} \hspace{1cm} (3)$$

Update the best sequence and $Z_{best}$, if the generated sequence is superior to the best sequence that has been obtained so far.

4.2. Variable Neighborhood Search

This method is different from the most local search heuristics in that it uses two or more neighborhoods, instead of one, in its structure. In addition, to avoid costing too much computational time, the best number of neighborhoods is often three [16], which is followed by our algorithm, so index $l$ is defined to show local search type. The three neighborhoods employed in our algorithm are defined below:

1. Job swaps on one machine: one machine is chosen and all possible job swaps are considered.
2. Job swaps between two different machines: two machines are chosen in one stage and all possible job swaps from different machines are considered.
3. Job transfers from one machine to another: one machine is chosen and all possible job movements from this machine to any other are considered.
Note that the neighborhoods themselves \((N_i(S), N_2(S)\text{ and } N_3(S), \text{ respectively})\) are determined both by its respective structure and by the solution it is being applied to. The size of neighborhood \(N_i(S)\) is \(O(mn^2)\), neighborhood \(N_2(S)\) is \(O(m^2n^2)\) and neighborhood \(N_3(S)\) is \(O(n^2)\).

VNS always tries to use the fastest local search available first. If after an iteration no improvement is made, then another neighborhood is used \((l\text{ is incremented})\), and every time a new solution is found, the first and the fastest local search is used \((l = 1)\).

### 4.3. Simulated Annealing

SA is inspired by the physical process of heating a substance and then cooling it slowly, until a strong crystalline structure is obtained. This process is simulated by lowering an initial temperature by slow stages until the system reaches to an equilibrium point, and no more changes occur. The main idea of this technique is to start from some initial solution, \(\pi', \text{ and successively move a among neighboring solutions until the stopping condition is satisfied. At each iteration, } \pi' \text{ is selected from the neighborhood of actual solution } (\pi_{iter}) \text{ and it replaces with a probability}

\[
P(t_{iter}, \pi_{iter}, \pi') = \min \left\{ 1, \exp\left(\frac{-f(\pi') - f(\pi_{iter})}{t_{iter}}\right) \right\}
\]

where \(t_{iter}\) is a parameter called the temperature at iteration \(iter\). The temperature decreases during the search process according to the cooling scheme. The performance of SA depends on the following parameters, which have to be precisely selected:

#### 4.3.1. Initial Temperature

The initial temperature is selected on the basis of \(K = nm + 1\) solutions \(\pi_0; \pi_1; \ldots; \pi_K\), where \(\pi_j\) is randomly selected from the neighborhood of \(\pi_{j-1}\) and \(\pi_0\) is an initial solution. The initial temperature is defined as

\[
t_0 = \frac{\delta}{mn},
\]

where \(\delta = \max_{j=1,\ldots,k} \{f(\pi_j) - f(\pi_{j-1})\}\).

#### 4.3.2. Cooling Scheme

The temperature changes in every iteration according to the logarithmic cooling scheme:

\[
t_{j+1} = \frac{t_j}{1 + \lambda t_j},
\]

where parameter \(\lambda\) is defined as

\[
\lambda = \frac{\ln(t_f) - \ln(t_j)}{f/2}.
\]

and \(f\) is the total number of iterations. The final temperature \(t_f\) is determined from the following expression:

\[
t_f = \frac{\phi}{nm}.
\]

### 4.4. A Hybrid Method

The basic proposed hybrid algorithm structure is designed as below.

**Hybrid metaheuristic pseudocode**

1: \(R_{time} \leftarrow \text{Set run time}()\)
2: while run time < \(R_{time}\) do
3: \(S^* \leftarrow \text{Generate initial solution}()\)
4: \(l \leftarrow 1;\)
5: for each ant do
6: \(S \leftarrow S^*;\)
7: \(k \leftarrow 0;\)
8: \(T_k \leftarrow \text{Set initial temperature}()\)
9: \(T_f \leftarrow \text{Set final temperature}()\)
10: while current temperature > \(T_f\) do
11: \(\text{Shake procedure: find a random solution } S' \in N_i(S);\)
12: \(\text{Perform a local search on } N_i(S')\text{ to find a solution } S'\);\)
13: if \(f(S') \leq f(S^*)\) then
14: \(S^* \leftarrow S'\);
15: \(l \leftarrow 1;\)
16: else
17: accept $S'$ as new solution with probability $p(T_k, S', S^*)$
18: end if
19: Adapt temperature($T_i$)
20: $l \leftarrow l + 1$
21: end while
22: generate schedule
23: evaluate schedule
24: end for
25: verify for global or local best
26: evaporate pheromone in all trials
27: deposit pheromone on best global schedule
28: end while

4.5. STOPPING RULE

The stopping criterion used when testing all instances with the algorithms is set to a computational time (CPU time) limit fixed to $(n^2 \times \sum_{t=1}^{g} m' / g) \times 3$ milliseconds. This stopping criterion is not only responsive to the number of stages, but also is sensitive towards the rise in number of jobs and number of machines at each stage.

5. EXPERIMENTAL DESIGN

In this section the proposed hybrid algorithm with the RKGA proposed by Kurz and Askin [13] and immune algorithm (IA) which proposed by Zandieh et al. [14] for the SDST hybrid flowshop scheduling are compared. These algorithms originally were proposed for makespan objective, so, these algorithms for ET objective problems are adapted. All algorithms are coded in C++ and run with an Intel Pentium IV dual core 2.5 GHz PC at 896 MB RAM under a Microsoft Windows XP environment.

5.1. DATA GENERATION AND SETTINGS

An experiment was conducted to test the performance of the hybrid algorithm. Following Kurz and Askin [12], data required for a problem consists of the number of jobs, number of stages, number of machines in each stage, range of processing times, and the range of sequence dependent setup times. The problem data can be characterized by five factors, and each of these factors can have at least two levels. These levels are listed in Table 1.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of jobs</td>
<td>6, 30, 100</td>
</tr>
<tr>
<td>Machine distribution</td>
<td>Constant: 1, Variable: Uniform(1, 4), Uniform(1, 10)</td>
</tr>
<tr>
<td>Number of stages</td>
<td>2, 4, 8</td>
</tr>
<tr>
<td>Processing times</td>
<td>Uniform(50, 70), Uniform(20, 100)</td>
</tr>
<tr>
<td>Skipping probability</td>
<td>0.00, 0.05, 0.40</td>
</tr>
</tbody>
</table>

In general, all combinations of these levels are tested. However, some further restrictions are introduced. The variable machine distribution factor requires that at least one stage has a different number of machines than the others. Also, the largest number of machines in a stage must be less than the number of jobs. Thus, the combination with 10 machines at each stage and 6 jobs will be skipped and the combination of 1–10 machines per stage with 6 jobs will be changed to 1–6 machines per stage with 6 jobs. There are 252 test scenarios and ten data sets are generated for each one.

Another important issue is the due dates of the jobs. To generate due dates of all $n$ jobs the following steps is proposed:

1. Compute total processing time of each job on all $g$ stages.

   \[ p_i = \sum_{j=1}^{g} p_{ij}, \forall i \in N \]  

   \[ (9) \]

2. Compute average setup time for all possible subsequent jobs and sum it for all $g$ stages.

   \[ s_i = \sum_{j=1}^{n} \sum_{k=1}^{n-j} \frac{S_{jk}'}{n-1}, \forall i \in N \]  

   \[ (10) \]

Determine a due date for each job.
\[ d_i = (1 + \text{random} \times 3) \times \left( \frac{g}{\min_{m' \in M} (m')} \right) \times (s_i + p_i), \; \forall i \in N \]  

where \text{random} is a random number from a uniform distribution over range \((0, 1)\). This method of generating due dates results in very tight to relatively tight due dates.

### 5.2. Hybrid Algorithm Parameters Tuning

The efficiency of a hybrid algorithm depends on the choice of the best parameters in order to prevent premature convergence, to ensure diversity in the search space, to intensify the search around interesting regions, etc. This section describes an empirical testing approach for finding the best tuning parameters of our hybrid algorithm. In order to reduce the number of combinations, some parameters of the HM are fixed a priori based on the claims suggested in the literature. Parameters tuning is applied only for the minimum values for trail intensity \(\tau_{\min}\) with two levels \(1/5 \tau_{\max}\) and \(1/10 \tau_{\max}\), persisting the trail intensities \(\rho\) with three levels \((0.8, 0.9\) and \(0.95)\) in ACO, initial temperature parameter \(\varphi_1\) with \(0.7, 0.9\) and \(\text{Uniform} (0.5, 1)\) and final temperature parameter \(\varphi_2\) with three levels \((0.01, 0.1\) and \(\text{Uniform} (0, 0.1)\)) in SA algorithm.

54 different hybrids are obtained by these levels. Six instances are generated randomly, 2 small, 2 medium and 2 large, for each combination of \(n, m, SDST\) resulting a total of 252 instances. All these instances are solved by 54 different hybrid algorithms. The results are analyzed by the means of multi-factor Analysis of Variance (ANOVA) technique. The three main hypotheses for the validity of ANOVA, i.e.: randomness and independence, normality, and homogeneity of variance have been tested. That was done and founded no bias for questioning the validity of the experiment (see Montgomery [17]). Table 2 shows the results for different sizes of problems; small, medium and large.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau_{\min})</td>
<td>(1/5 \tau_{\max})</td>
<td>(1/5 \tau_{\max})</td>
<td>(1/10 \tau_{\max})</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.9</td>
<td>0.95</td>
<td>0.9</td>
</tr>
<tr>
<td>(\varphi_1)</td>
<td>Uniform ((0.5, 1))</td>
<td>Uniform ((0.5, 1))</td>
<td>Uniform ((0.5, 1))</td>
</tr>
<tr>
<td>(\varphi_2)</td>
<td>Uniform ((0.01, 1))</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

### 5.3. Evaluation Metrics

After computation of \(\sum (E + T)\) of each algorithm for its instances, the best solution obtained for each instance (which is named \(\text{Min}_{\text{sol}}\)) by any of the three algorithms is calculated. Relative Percentage Deviation (RPD) is obtained by the following formula:

\[
\text{RPD} = \frac{\text{Alg}_{\text{sol}} - \text{Min}_{\text{sol}}}{\text{Min}_{\text{sol}}} 
\]  

where \(\text{Alg}_{\text{sol}}\) is the \(\sum (E + T)\) obtained for a given algorithm and instance.

### 5.4. Numerical Result

The results of the experiments for three subsets, averaged for each one of the \(n\) and \(g\) configurations are shown in Table 3. Each instance is solved using 6 different seeds and the average solution is considered.

As it can be seen, the hybrid algorithm provides better results than RKGA and IA. In order to verify the statistical validity of the results shown in Table 3 a design to confirm which the best algorithm is, a design of experiments and an analysis of variance (ANOVA) is performed where the different algorithms are considered as a factor and the response variable RPD.

<table>
<thead>
<tr>
<th>Problem size</th>
<th>RKGA_\text{sol}</th>
<th>IA_\text{Zandieh}</th>
<th>HMH</th>
</tr>
</thead>
<tbody>
<tr>
<td>6\times2</td>
<td>0.2254</td>
<td>0.2122</td>
<td>0.0296</td>
</tr>
<tr>
<td>6\times4</td>
<td>0.2191</td>
<td>0.2017</td>
<td>0.0422</td>
</tr>
<tr>
<td>6\times8</td>
<td>0.2183</td>
<td>0.2014</td>
<td>0.0326</td>
</tr>
<tr>
<td>6 Job</td>
<td>0.2185</td>
<td>0.2051</td>
<td>0.0348</td>
</tr>
<tr>
<td>30\times2</td>
<td>0.1773</td>
<td>0.1528</td>
<td>0.0453</td>
</tr>
<tr>
<td>30\times4</td>
<td>0.3921</td>
<td>0.2371</td>
<td>0.0968</td>
</tr>
<tr>
<td>30\times8</td>
<td>0.2921</td>
<td>0.1817</td>
<td>0.0487</td>
</tr>
<tr>
<td>30 Job</td>
<td>0.2872</td>
<td>0.1905</td>
<td>0.0636</td>
</tr>
<tr>
<td>100\times2</td>
<td>0.2132</td>
<td>0.1574</td>
<td>0.0369</td>
</tr>
</tbody>
</table>
5.5. **ANALYSIS OF CONTROLLED FACTORS**

**ANALYSIS OF PROBLEM SIZE FACTOR (NUMBER OF JOBS)** In order to see the effect of number of jobs on three algorithms, a two ways ANOVA is applied. Means plot and least significant difference (LSD) intervals (at the 95% confidence level) for the interaction between the factors of type of algorithm and number of jobs are shown in Figure 1.

![Figure 1: Plot of RPD for the interaction between the type of algorithm and number of jobs](image)

As it could be seen, in the all cases of \( n = 6, n = 30 \) and \( n = 100 \) the HMH works better than RKGA and IA.

**ANALYSIS OF G FACTOR (NUMBER OF STAGES)** Another two ways ANOVA and LSD test are applied to see the effect of magnitude of stages on quality of the algorithms. The results are shown in Figure 2. As it could be seen, in all cases the hybrid algorithm works better than RKGA and IA.

**ANALYSIS OF M FACTOR (NUMBER OF MACHINES)** Another two ways ANOVA and LSD test are applied to see the effect of magnitude of machines on quality of the algorithms. The results are illustrated in Figure 3. As it could be seen, in all cases the hybrid algorithm works better than RKGA and IA.

6. **SUMMERY**

This study has addressed the problem of scheduling in the hybrid flowshop with sequence-dependent setup time. For the objective of sum of the earliness and tardiness about a due date of jobs, the hybrid metaheuristic algorithm which combines ACO, SA and VNS in a population-based context is suggested. For each new generation of schedules, the key is to use hybridizing the population-based evolutionary searching ability of ACO with the local improvement ability of some VNS and SA to balance exploration and exploitation. In the proposed HMH, the balance between the global exploration and the local exploitation was stressed. This approach generated new individuals that can effectively guide and speed up the search.

When tested on many problem instances, this model proved to be efficient and robust. The computational results for test problems with different structures indicate that the HMH provides good average results for many test problems compared to the genetic algorithm and immune algorithm proposed in a lately published.

We believe that the contribution of this method is not only the development of a framework by which due-date-related problems can be solved, but also the generality of the framework, as it can solve different categories of scheduling problems readily and with very good performance. For future work, this research can be extended in several directions. For example, this study can be extended to consider other versions of the problems; e.g., single machine or job shop scheduling problems, since general ideas of the algorithm can be easily
implemented in various scheduling problems.

REFERENCES

Setting warm-up period in simulation analysis with autocorrelated data

S.M.T.Fatemi Ghomi * and S.Shokuhyar

Department of Industrial and Engineering, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, Iran

ABSTRACT

One means for dealing with initialization bias in simulation experiments is to implement a warm-up period. This requires the correct estimation of the initial transient. In literature many methods has been used to determine the warm-up period for non autocorrelated data. But simulation output data are typically auto correlated. So in this paper auto correlated time series model is applied to select appropriate time series. For doing this, one software is develop to generate partial autocorrelation function and sample autocorrelation function in simulation software to set appropriate parameter for time series. After selecting appropriate time series, control limit for system is set. Then with running simulation model, warm up period will be defined. To apply this method, at least one example is presented.

1. INTRODUCTION

Most simulation runs are started with empty systems. For example, a production line is usually started with no parts in any of the buffers or machines. Since there is no congestion at this time, parts entering such an empty system tend to flow through the system at a faster rate than parts produced during more representative hours. Including information about these early parts will cause to under-estimate the value of most of steady state performance measures such as the average production rate or the mean time spent by parts in the system. This is illustrated in the following Figure1 of the fraction of parts spending more than 60 minutes in a three station flow line for five different simulation runs. Each point represents the average value computed for all parts having left the system at the indicated time. The plot includes all averages for the first 1,000 customers.

Figure 1: Plot of the first 1000 values of the fraction of parts spending more than 60 minutes in a three station flow line for five different simulation runs

It is seen that the observed average fraction always starts out at zero, then increases and eventually converges to

*Corresponding Author :Tel.: (+9821)66413034; Fax: (+9821)66413025; Email address: Fatemi@aut.ac.ir
about 60% - 80%. However, the effect of this initial bias will never be completely eliminated. Also it is seen that each run appears to be quite noisy, but that the average value for each observation appears to be better behaved.

The effect of the initial bias can be eliminated if data collection is not started until the simulation has reached a more representative "steady state". For example, it appears that the bias would be eliminated in the above if the first 300 observations were eliminated [1].

The goal of steady state simulation studies is to estimate the long run characteristics of the system. In order to estimate the long run performance measures, the simulation model is run for a certain period of time. Most of the simulation models are started empty and idle. Almost every time, these conditions differ from the steady state condition. Due to this, the simulation model takes some time to reach steady state. During this time period the model is said to be in transient state. The observations collected during this time period may affect the accuracy of the estimates of the performance measure, if the transient state lasts for a relatively long time. This problem is called as the initialization bias or the startup problem in simulation literature. One of the ways to overcome this problem is to run the simulation model for a time period L called as the warm-up length. Reset all statistics after time L and start recording observations for a period of m - L, where m is the run length. Another difficulty here is in predicting the warm-up length [2].

So the practical question is "How long should the warm-up period be?"

Regarding the above discussion in estimating the steady-state parameters of a simulation model, the correct removal of any initialization bias is of vital importance. There are two methods open to the simulation modeler.

1) - the starting condition of the model can be set such that no bias exists in the output data. This requires the correct setting of the starting condition.
2) - the model can be run for a warm-up period and the data are then deleted from that period (the initial transient). This, of course, requires the correct estimation of the warm-up period. This paper focuses on this latter approach.

The purpose of this paper is to describe a new method to estimate the warm-up period of a steady-state simulation. Before describing this method there is a brief review of the literature on estimating the warm-up period. Then the new method is described. To apply this method, at least one example is presented.

2.Methods to estimate the warm-up period

Various methods have been purposed to identify the initial transient in simulation models. The devised warm-up methods can be categorized under five headings [3]:

- **Graphical methods:**
  These involve the visual inspection of time-series of the output data; such as Welch’s Method [4]. This is the simplest and most general technique used to determine the warm-up length. It is a graphical technique that requires multiple replications.

- **Heuristics approaches:**
  These apply simple rules, with few underlying assumptions.

- **Statistical methods:**
  These rely upon the principles of statistics for determining the warm-up period; such the method used by Yücesan [5]. He presented a method to detect the initialization bias. The method is based on randomization tests. Yücesan formulated the problem of initialization bias in a hypothesis testing framework concerning the mean of the process. Randomization tests are applied to test the null hypothesis that mean of the process is unchanged throughout the run. The advantage of using this method is that, no assumptions, like that of normality are required. The null hypothesis for this method is that there is no initialization bias.

- **Initialization bias tests:**
  These identify whether there is any initialization bias in the data and, therefore, they are not strictly methods for identifying the warm-up period, but they can be used in combination with warm-up methods to determine whether they are working effectively.

- **Hybrid methods:**
  These involve a combination of graphical or heuristic methods with an initialization bias test.

One of the most useful methods to determine the warm-up period is Statistical Process Control (SPC) method to detect when a model is in steady-state [3].

There is a close relationship between the concepts of SPC and those of transient and steady-state in simulation
output analysis. A model that is in steady-state varies about a constant mean according to some fixed distribution, while in transient it does not. As such, a simulation that is in steady-state could be considered to be “in-control”, while during a transient phase it could be considered to be “out of control”. It seems possible; therefore, that the methods employed in SPC could be used to detect when a model is in steady-state, and when it is not [3].

But the main weakness of this method, is that the process requires to be non autocorrelated. Simulation output data are typically autocorrelated and so SPC methods need some assumptions for applying to determine the warm-up period. So in the next section autocorrelated time series model is applied to determine the warm-up period.

3. The proposed model

Conventional control charts are based on the assumption that the observations are independently and identically distributed (i.i.d.) over time.

Numerous control chart modifications have been proposed for monitoring auto correlated processes. But three views can be considered for applying auto correlation in time series data [6]:

1. At one extreme, Wheeler [7] argues that the usual control limits are contaminated “only when the autocorrelation becomes excessive (say 0.80 or larger).” He concludes that “one need not be overly concerned about the effects of autocorrelation upon the control chart.”

2. At the opposite extreme, automatic process control (APC), also referred to as engineering process control, views autocorrelation as a phenomenon to be exploited. In contrast to SPC, which assumes that the process remains on target unless an unexpected but removable cause occurs, APC assumes that the process is changing dynamically due to known causes that cannot be eliminated. Instead of avoiding “over control” and “tampering,” which have a negative connotation in the SPC framework, APC advocates continuous tuning of the process to achieve minimum variance control. Descriptions of this approach and discussion of the differences between APC and SPC are provided by a number of authors, including Box and Kramer [8], Macgregor [9,10], Macgregor, Hunter, and Harris [11], and Montgomery and others [12].

3. A third strategy advocates removing autocorrelation from the data and constructing a Shewhart chart (or an EWMA chart or a CUSUM chart) for the residuals; refer, for example, to Alwan and Roberts (1988).

In this paper, process, $X_t$, is considered an ARMA (1, 1) model, because most auto correlated manufacturing processes can be modeled by the popular first-order ARMA model [13].

As with the process dynamic model, the ARMA model can be represented in several forms. Here the state-space form is considered which will be used in the simulation modeling. The interested reader may refer to Box et al. [14].

for an introduction to state-space equations for ARMA models:

$$X_t = (1 - \phi_1 \mu_0 + \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

where $X_t$ is the observation at time $t = 1, 2, \ldots$, $\epsilon_t$ is the random noise at time $t = 1, 2, \ldots$ assumed to be i.i.d. normal $(0, \sigma_\epsilon)$, $\phi$ is the autoregressive parameter, $\theta$ is the moving average parameter and $\mu_0$ is the nominal mean of the process. The ARMA parameters, $\phi$, $\theta$ and $\mu_0$ are assumed to be known.

The following Sections should be taken to build the model. Section 3.1 describes setting simulation initial data and running simulation model. Section 3.2 introduces gathering collected data and set appropriate time series model. Section 3.3 and 3.4 deals with setting control limit for data and identify the initial transient.

3.1 Set simulation initial data and run simulation model

First, time-series data on a key output statistic need to be collected. Typically this would be throughput for manufacturing systems and customer waiting time in service systems. The output data should be collected over a series of replications to provide a number of samples for each data point. The observation interval should be relatively short; for typical manufacturing and service system models a one hour interval is probably appropriate.

The length of each replication also needs to be determined. It is recommended that the run length is at least four times longer than the estimated length of the initial transient [3].

Generally, however, identification and estimation of ARMA models should be based on at least 50 and preferably 100 or more observations [3] and at least 60 data points are collected in each replication. An initial estimate of the
length of the initial transient could be obtained by simply inspecting a time-series of the output data to look for where the model appears to be in steady-state.

For running simulation model Arena software is applied that has more capability for running simulation model [15].

### 3.2 GATHERING COLLECTED DATA AND SETTING APPROPRIATE TIME SERIES MODEL

After running simulation model, formulation of partial autocorrelation function and sample autocorrelation function are programmed in Arena software with visual basic language to determine type of autocorrelated time series model. The partial auto correlation function is defined as follows [13]:

\[
 r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}
\]

where \( r_k \) is partial autocorrelation at lag \( k \), \( N \) is number of data and \( \bar{x} \) is the average of data. Also sample autocorrelation function is described in Box & Jenkins [13]. After selecting appropriate time series model and setting parameters for this model, control limit is set for data.

As mentioned before, in this paper process, \( X_t \), is consider an ARMA(1, 1) model, because most auto correlated manufacturing processes can be modeled by the popular first-order ARMA model [13].

### 3.3 SETTING CONTROL LIMIT FOR DATA

SPC is traditionally applied to processes in which successive observations are assumed to be independent and identically distributed (i.i.d.). Unfortunately, the i.i.d. assumption is often violated in practice, and the presence of autocorrelation has a serious impact on the performance of the control charts. To deal with autocorrelated data, various charting techniques have been proposed. The residual-based chart or the special cause chart (SCC) is one of the most widely used methods [16, 17]. The key idea of the residual-based charts is to fit a time series model to subtract the autocorrelation and then to monitor the residuals. Assuming that the time series model is accurate, the residuals are statistically independent, and thus conventional control charts, such as the Shewhart, cumulative sum (CUSUM) or EWMA chart, can be applied to the residuals. The residual behavior of an autocorrelated process has gained extensive attention. Moskowitz et al. [17] studied the residual-based charting performance for ARMA (1, 1) process, and VanderWiel studied it for IMA (0, 1, 1) process. Hu and Roan [18] investigated the change patterns of the residuals for ARMA (1, 1) and ARMA (2, 1) processes. The relationship between the standard deviation, \( \sigma_x \), of the ARMA (1, 1) process, \( X_t \), and the standard deviation, \( \sigma_\varepsilon \), of the random noise, \( \varepsilon_t \), is known to be[19]:

\[
 \sigma^2_x = \frac{1 + \theta^2 - 2\phi\theta}{1 - \phi^2} \cdot \sigma^2_\varepsilon
\]

(3)

Here, the SCC is merely an individual chart for the random noise, \( \varepsilon_t \), i.e:

\[
 \varepsilon_t = X_t - (1 - \phi)\mu_0 - \phi X_{t-1} - \theta \varepsilon_{t-1}
\]

(4)

where \( \varepsilon_t = X_t - \mu_0 \); since \( \varepsilon_t \) is assumed to be an i.i.d. normal \( (0, \sigma_\varepsilon) \) random variable when the process is in-control. In this situation the control limits of the SCC chart are simply:

\[
 CL = \pm K \sigma_\varepsilon
\]

(5)

where \( K \) is a constant that is usually chosen to be 3 (see Montgomery [20]).
3.4 **IDENTIFY THE INITIAL TRANSIENT**

The final step is to view the plot of the control chart. This step is the same as Robinson method [3] during the initial transient the time-series data will be “out of control”. Once the model reaches steady-state, the data will be “in-control”. Bissell identifies a number of rules to determine whether a control chart is showing data that are out of control [21]:

- The time-series data violates an action limit (note that occasional violations, 1 in 1000, are to be expected for in-control data, and can be explained by Type I errors).
- Two consecutive values violate either the upper or the lower warning limit.
- Frequent values in relatively close succession that violate a warning limit.
- Persistent trend in the time-series data.
- A run of seven or more values on either side of the central line.
- Excessive zigzagging, with few points near to the central line and many near to the control limits.

The warm-up period for the model can be selected by identifying the point at which the time-series data is in-control and remains in-control.

4. **EXAMPLE FOR APPLICATION OF PROPOSED METHOD**

In order to implement the application of method, one simulation example is demonstrated. These data are used as output data to simulate an ARMA (1, 1) process.

Ten replications were performed with a run length of 1,000 hours. Statistics were collected on hourly throughput consisting of 100 observations, $X_t$, that correspond to an in-control process. In Figure 2, plot for the classical $3\sigma$ control limits for the SCC chart is showed, with $\pm 3\sigma = \pm 0.352$.

![Figure 2: Output analysis of simulation model with control limits](image-url)

As shown in Figure 2, if the process is assumed in-control, 3 points are out of control limits. So in this example 25 time unit is appropriate for warm-up period.
5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

In the previous related works, simulation output is supposed to be none correlated. But typically because of
simulation nature, output is autocorrelated and so these methods were not appropriate to determine warm-up period
and get not proper result. So, this paper studies warm-up period in discreet event simulation and develops algorithm
to determine the warm-up period with autocorrelated data. For doing this Box & Jenkins model is applied. One
example is given to apply the developed algorithm and showed that this algorithm how work in sample problem.

The following areas of research are recommended for future studies:
1. Other types of autocorrelated time series model such as ARIMA, AIMA and etc models can be
   considered in future studies.
2. Other types of control limits can be taken in the future research.
3. Some new studies are made to automation of warm-up period detection by applying some
   programming code in simulation software for determining optimized autocorolated model.

ACKNOWLEDGEMENTS

The authors would like to express sincere thanks to anonymous referees for valuable comments which greatly
improved the presentation of this paper.

REFERENCES

[6] Sas online doc, version eight, chap49, can be accessed: http://www2.stat.unibo.it/ManualiSas/qc/chap49.pdf
[9] MacGregor, J., “Interfaces between process control and online statistical process control”, computing and systems
[14] Nemhhard, HB,” Simulation using the state-space representation of noisy dynamic systems to determine effective
Setting warm-up period in simulation analysis with autocorrelated data


Efficient Stochastic Hybrid Heuristics for the Multi Depot Vehicle Routing Problem

M. Mirabi, S. M. T. Fatemi Ghomi, F. Jolai

ACECR Higher Education Institute, Yazd, Iran

Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, Iran

Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

ABSTRACT

The paper addresses the problem of multi depot vehicle routing in order to minimize the delivery time of vehicle objective. Three hybrid heuristics are presented to solve multi-depot vehicle routing problem. Each hybrid heuristic combines elements from both constructive heuristic search and an improvement technique. The improvement techniques are deterministic, stochastic and simulated annealing (SA) methods. Experiments have been run on a number of randomly generated test problems of varying depots and customer sizes. Our heuristics are shown to outperform one of the best-known existing heuristic. Statistical tests of significance are performed to substantiate the claims of improvement.

1. INTRODUCTION

The vehicle routing problem (VRP) is a generic name referring to a class of combinatorial optimization problems in which customers are to be served by a number of vehicles. The vehicles leave the depot, serve customers in the network and on completion of their routes return to the depot. Each customer is described by a certain demand. Other information includes the coordination and distance between depot and customers and the capacity of the vehicles providing the service. In the other word, VRP involves the design of a set of routes for a fleet of vehicles, starting and ending at a depot and serving a number of customers with known demands. Each customer must be visited by one of these routes and the objective is to minimize the global cost of the set of routes [1]. The VRP has extensive variants, including the PVRP: periodic VRP in which the customers are served in a period of time rather than one day, the VRPPD: VRP with pickup and delivery in which the customers may both receive and send products, the VRPTW: VRP with time windows in which the vehicles must arrive at the customers before the latest arrival time, while arriving before the earliest arrival time results in waiting and so on [2]. A common point of the above versions is that they are all based on one depot. Thus, they can be regarded as single-depot VRPs. Although the single-depot VRPs have attracted so much attention, they are not suitable for some cases where a company has more than one depot. Due to this reason, this paper focuses on the single-depot and multi-depot VRP (MDVRP) in which more than one depot is considered. Because there are additional depots for storing the products, the decision makers also have to determine which customers are served by which depots, that is, the grouping problem prior to the routing and scheduling problems. Obviously, this type of problem is more challenging and sophisticated than the single-depot VRPs. Besides, the MDVRP is NP-hard, which means that an efficient algorithm to solve the problem to optimality is unavailable. Therefore, solving the problem by an exact algorithm is time consuming and computationally intractable. To deal with the problem efficiently and effectively, three probabilistic hybrid heuristics are developed in this paper.

This paper is organized as follows. Section 2 surveys on the relevant literature. Section 3 describes the hierarchy of decisions in the MDVRP. Section 4 discusses the principles of the algorithms used to solve the MDVRP. Section 5 compares the performance of the algorithms. Finally, Section 6 concludes the paper.

2. LITERATURE REVIEW

Over the last 30 years or so, the classical VRP has attended strongly in the literature. Comparatively, the number of research projects on the MDVRP is fewer. Tabu search heuristics for the MDVRP have been proposed by Renaud
et al. [3] and Cordeau et al. [4]. The recent paper deserves special attention as it describes a general heuristic that also solves periodic vehicle routing problems (PVRP) and periodic traveling salesman problems. Earlier, Chao et al. [5] proposed a record improvement heuristic for the MDVRP.

Ho et al. [2] developed two hybrid genetic algorithms (HGAs) for MDVRP. The major difference between the HGAs was the initial solutions. Pisinger and Ropke [6] presented a unified heuristic which is able to solve different variants of the vehicle routing problem. Crevier et al. [7] addressed an extension of the multi-depot vehicle routing problem that vehicles may be replenished at intermediate depots along their routes. Salhi and Sari [8] proposed a heuristic method with three levels to solve the MDVRP. The first level was the construction of an initial feasible solution. The second and the third levels were to improve the routes in each depot, that is, intra-depot and the routes in all depots, that is, inter-depot, respectively. Su [9] proposed a dynamic vehicle control and scheduling system to solve the MDVRP. All the control decisions were made according to the real time status of the system, such as the location, quantity, and due date of the demand. Wu et al. [10] studied the multi-depot location-routing problem (MDLRP), which is an extension of the MDVRP. The MDLRP was decomposed into the location-allocation problem and the VRP, and then they were solved sequentially and iteratively using the simulated annealing (SA). The major difference between the MDLRP and the MDVRP is that the former also determined the number and locations of depots. Giosa et al. [11] investigated the multi-depot VRPTW (MDVRPTW), which is also an extension of the MDVRP. The authors designed and compared six heuristics for assigning the customers to depots while using the same VRP heuristic for each depot. Similar to Wu et al. [10], Wasner and Zapfel [12] also studied the MDLRP for planning of parcel service. A heuristic method based on the local search with a series of feedback loops was developed to solve the problem separately. Nagy and Salhi [13] presented a number of heuristic methods to solve the single-depot VRPPD. The methods can be modified to tackle the multi-depot VRPPD (MDVRPPD).

Due to the complexity of the problem, solving the MDVRP to optimality is extremely time-consuming. To tackle the problem efficiently, all previous researchers preferred heuristic methods to exact algorithms. According to the above literature review, there are two common points among these proposed methodologies. First, the MDVRP was decomposed, and then the sub problems were solved sequentially and iteratively. Second, the heuristic methods consisted of two mechanisms: construction and improvement. The first mechanism generated initial feasible solutions, whereas the second mechanism modified the existing solutions to yield better results. However, it is found that none of the researchers have applied the probabilistic hybrid heuristics to solve the MDVRP. This is our primary motivation for writing this paper.

3. MULTI-DEPOT VEHICLE ROUTING PROBLEM

Consider a distribution company with multiple depots. The number and locations of the depots are predetermined. Each depot is large enough to store all the products ordered by the customers. A fleet of vehicles with limited capacity is used to transport the products from depots to customers. Each vehicle starts and finishes at the same depot. The location and demand of each customer is also known in advance. Each customer is visited by a vehicle exactly once. This practical distribution problem can be regarded as the MDVRP, in which there are three decisions as: 1) grouping: assigning customers to depots, 2) routing: assigning customers in each depot to routes and 3) scheduling: sequencing each route in every depot. The decision makers first need to cluster a set of customers to be served by the same depot, that is, the grouping problem. Then they have to assign customers of the same depot to several routes so that the vehicle capacity constraint is not violated. At last, the decision on delivery sequence of each route is made. Generally, the objective of the MDVRP is to minimize the total delivery distance or time spent in serving all customers. Shorter delivery time results in higher level of customer satisfaction.

Besides, the objective can also be the minimization of the number of vehicles required. Fewer vehicles imply that the total operation cost is reduced. No matter which type of objectives is defined, the ultimate goal of the MDVRP is to increase the efficiency of the delivery.

3.1. FORMULATION

The objective is to design the delivery routes from a depot to customers, i.e., to determine a sequence of customers on each vehicle route. Customer demands are deterministic and known. The time required to travel between customers including the depot is deterministic. All the vehicles have the same capacity. Each vehicle starts the travel from a depot and has to return to a depot in completion of services to customers. The notations and mathematical formulation are as follows:
Sets

- $I$: Set of all depots
- $J$: Set of all customers
- $K$: Set of all vehicles

Parameters

- $N$: Number of customers
- $C_{ij}$: Distance between point $i$ and $j$, $i, j \in I \cup J$
- $V_I$: Maximum throughput at depot $i$
- $d_j$: Demand of customer $j$
- $Q_k$: Capacity of vehicle (route) $k$

Decision variables

- $x_{ik}$: 1, if point $i$ immediately precedes point $j$ on route $k$ ($i, j \in I \cup J$); 0, otherwise.
- $z_j$: 1, if customer $j$ is allocated to depot $i$; 0 otherwise.
- $U_a$: auxiliary variable for sub-tour elimination constraints in route $k$.

Mathematical model:

\[
\begin{align*}
\text{Min} & \quad \sum_{i \in I} \sum_{j \in I \cup J} \sum_{k \in K} C_{ij} x_{ik} & (1) \\
\sum_{k \in K} x_{ik} & = 1, \quad j \in J & (2) \\
\sum_{k \in K} x_{jk} & = 1, \quad i \in I & (6) \\
\sum_{j \in J} d_j x_{jk} & \leq Q_k, \quad k \in K & (3) \\
\sum_{k \in K} x_{ik} - \sum_{k \in K} x_{jk} & = 0, \quad k \in K, i \in I \cup J & (5) \\
\sum_{j \in J} x_{jk} & \leq 1, \quad k \in K & (7) \\
\sum_{j \in J} z_j & \leq V_I, \quad i \in I & (7) \\
\sum_{i \in I} z_j & + \sum_{j \in J} (x_{ij} + x_{ji}) & \leq 1, \quad i \in I, j \in J, k \in K & (8) \\
x_{ij} & \in \{0,1\}, \quad i \in I, j \in J, k \in K & (9) \\
z_j & \in \{0,1\}, \quad i \in I, j \in J & (10) \\
U_a & \geq 0, \quad l \in J, k \in K & (11)
\end{align*}
\]
The objective function minimizes the total delivery distance. Eqs. (2) require that each customer be assigned to a single route. Eqs. (3) are the capacity constraint set for vehicles. Eqs. (4) are the new sub-tour elimination constraint set. Flow conservation constraints are expressed in (5). Constraints (6) assure that each route can be served at most once. Capacity constraints for the depots are given in (7). Constraints (8) specify that a customer can be assigned to a depot only if there is a route from that depot going through that customer. Constraint sets (9) and (10) are the binary requirements on the decision variables. The $u_a$'s auxiliary variables taking positive values are declared in (11).

4. SIMULATED ANNEALING

Since the third proposed hybrid heuristic benefits simulated annealing (SA) method in the improvement phase, following gives a short explanation of this method.

SA of Aarts and Korst [14] and Van Laarhoven and Aarts [15], a heuristic search method based on ideas drawn from statistical physics, has been found to be effective in many combinatorial optimization problems. The algorithm begins with a randomly generated initial point (the trial solution). This initial solution is the “current” solution. A neighbor (an adjacent point) of this current solution is then generated, following some predetermined neighbor-generating method. If the neighbor is found to be better than the current point, it (the neighbor) is unconditionally accepted as the new current point. On the other hand, if the neighbor is found to be worse, it is not rejected outright, but accepted with a certain probability. The algorithm proceeds by iterating a certain number of times over the transition from the current point to the adjacent point (which becomes the next current point). At the beginning of a SA run, the probability of accepting a worse point is kept high (thereby reducing the chance of the SA algorithm getting trapped in a local optimum). As the number of iteration increases, this probability is reduced according to a specific policy. It is customary to use a control parameter, called the temperature (analogous to the temperature of the physical process), to alter the probability of acceptance/rejection. Usually, the temperature is started at a high value and is gradually brought down according to a schedule known as the annealing (cooling) schedule. This annealing schedule determines how the probability of accepting a worse point decreases with time (iterations).

5. THE PROPOSED HYBRID HEURISTICS

Each hybrid heuristic methods combines elements from both constritive heuristic search and a improvement technique. The improvement techniques are deterministic, stochastic and simulated annealing methods. Now based on the above explanations, following contains three hybrid heuristic methods.

5.1. INITIALIZATION

The path representation is used to encode the solution of the MDVRP. The idea of the path representation is that the customers are listed in the order in which they are visited. For example, suppose that there are six customers numbered 1–6. If the path representation is (0 2 4 1 0 3 6 5 0), then two routes are required to serve all these six customers. In the first route, a vehicle starts from the depot, which is denoted as 0, travels to customers 2, 4, and finally customer 1. After that, the vehicle returns back to the depot. In the second route, the vehicle starts with customer 3, then customer 6, and finally customer 5. Similarly, the vehicle travels back to the depot after serving the customers. Note that each chromosome contains n links if there are n depots in the MDVRP. In the stage of initialization, there are three steps to generate a feasible initial solution. The first step is to assign customers to each of n links, that is, the grouping problem.

There are a number depots and customers and each customer must be assigned to one depot or link. Because the objective here is to minimize the total delivery time spent in distribution, customers are assigned to the nearest depot. For example, there are two depots, that is, $d_s$ and $d_a$ available in the MDVRP. Each customer, say $c_i$, should be assigned to a single depot exactly. The selection is based on the following calculation:

- If $D(c_i,d_s) < D(c_i,d_a)$, assign $c_i$ to $d_s$;
- If $D(c_i,d_s) > D(c_i,d_a)$, assign $c_i$ to $d_a$;
- If $D(c_i,d_s) = D(c_i,d_a)$, select the depot arbitrarily;

where $D(c_i,d) = \sqrt{(x_c - x_d)^2 + (y_c - y_d)^2}$.
represents the distance between customer i and depot k.

The second step is to assign customers in the same link to several routes using the Clarke and Wright [16] saving method. The saving regarded in this paper is the distance traveled by the vehicles to serve the customers. The method is to construct a saving matrix, $S(i, c, k)$, for every two customers in the same link first. Then, the customers with larger saving value are grouped in the same route while not violating the vehicle capacity constraint. We have

- **Saving matrix for Link 1:** $S(i, c, k) = D(d_i, c) + D(d_c, k) - D(d_i, c, k)$;
- **Saving matrix for Link 2:** $S(i, c, k) = D(d_i, c) + D(d_c, k) - D(c, k)$.

The third step is to solve the scheduling problem by the NNH [17]. The principle of the NNH is to start with the first customer randomly. Then the next customer is selected as close as possible to the previous one from those unselected customers to form the delivery sequence until all customers are selected.

### 5.2. IMPROVEMENT

Generally, the objective of the MDVRP is to minimize the total delivery distance or time spent in serving all customers. Shorter delivery time results in higher level of customer satisfaction. The corresponding objective function is represented as OB in all heuristics.

**Improvement M1**

1. For $i = 1$ to $n$ (number of customers) do
   - Randomly interchange any two customers belonging to the one route or two different routes in the solution obtained in the initialization step, find the OB of the modified schedule, and accept it as the current best solution if its OB is better, otherwise continue working with the original solution.
2. Use the output of the current solution as the final solution.

   The M1 may interchange the customers within the same route, that is, intra-route improvement or within the same depot, that is, intra-depot improvement. The M1 may also swap a customer from one route to another route, that is, inter-route improvement or from one depot to another depot, that is, inter-depot improvement. The type of improvements performed is dependent on the selection of customers in Step 1.

**Improvement M2**

Algorithm M2 attempts to improve upon M1 by deterministically accepting the adjacent solution if it is better than the current one, and accepting it probabilistically if it is not.

Here are the following steps:
1. For $i = 1$ to $n$ do
   - Randomly interchange any two customers belonging to the one route or two different routes in the solution obtained in the initialization step, find the OB of the modified schedule, and accept it as the current best solution if its OB is better, otherwise accept it with probability $e^{-\Delta/T}$, where $\Delta = OB(modified) - OB(original)$, and $T$ is a parameter of the algorithm that is held fixed at a predetermined value during the course of a run.
2. Use the output of the current solution as the final solution.

**Improvement M3**

This heuristic applies SA to improve upon the best solution obtained at any step of the algorithm.

The algorithm is described as follows:
1. For $i = 1$ to $n$ do
   a) Initialize Max-iterations, Temp-start.
      Set Count = 1, $T = Temp-start$.
      Let the best solution obtained in the initialization step be called the current solution, $x_0$.
      Compute OB($x_0$).
   b) Randomly generate a neighboring solution using either the interchange neighborhood, Forward Insertion Neighborhood or Backward Insertion Neighborhood (these neighborhoods are explained below). Let the neighboring solution be called the adjacent solution, $x_a$. Compute OB($x_a$).
   c) If $OB(x_a) < OB(x_i)$
      Then set $x_i = x_a$;
      Else
• Set $D = \text{OB}(x_i) - \text{OB}(x_j)$;
• Set $T = \text{Temp-start}/\log(1+\text{Count})$;
• With probability $e^{-\lambda/T}$ set $x_i = x_j$.

d) Increment Count by 1;
If Count $<$ Max-iterations, go to step b.

2. Use the output of the current solution as the final solution.

The annealing schedule used in step 1(c) of the above algorithm is due to Hajek [18]. The interchange neighborhood, by far the most popular scheme, is extremely simple: swap two randomly chosen customers in the sequence. Two other neighborhoods are something that Gupta and Smith [19] have introduced. In forward insertion neighborhood a customer is relocated further forward in the sequence and in backward insertion neighborhood a customer is relocated further backward in the sequence.

Three proposed methods are called by HM1, HM2 and HM3 respectively.

6. COMPUTATIONAL EXPERIMENTS

In this section, a computational study is carried out to compare the HM1, HM2, HM3 and the best assignment algorithm developed by Giosa et al. [11]. The performance of methods is evaluated using six randomly generated examples: 20-customer, 50-customer, 100-customer and 200-customer MDVRPs, in each of which there are two, three or five depots available. For each problem size, 10 independent problem instances were created. In Tables 1–3 each test suite comprises 10 independent instances. The OB values appearing in Tables are averages of 10 OB obtained from as many problem instances. The results indicate that HM1 and HM2 produce solutions of a better quality than assignment algorithm cited above for all problem sizes except two cases. Heuristic HM3 is vastly superior to HM1 or HM2 and outperforms assignment algorithm in all cases.

<table>
<thead>
<tr>
<th>Ins.</th>
<th>Number of depots</th>
<th>Number of customers</th>
<th>Ave. OB or $\overline{X}$</th>
<th>Ave. SD or $\overline{S}$</th>
<th>$T$</th>
<th>$\nu$</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>80.31</td>
<td>78.79</td>
<td>1.10</td>
<td>1.69</td>
<td>0.83</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>50</td>
<td>110.83</td>
<td>109.21</td>
<td>2.64</td>
<td>2.22</td>
<td>1.49</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>100</td>
<td>158.99</td>
<td>160.27</td>
<td>3.40</td>
<td>2.76</td>
<td>-0.92</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>200</td>
<td>221.74</td>
<td>222.84</td>
<td>3.86</td>
<td>3.21</td>
<td>-0.69</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>20</td>
<td>68.28</td>
<td>66.82</td>
<td>1.35</td>
<td>1.69</td>
<td>2.13</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>50</td>
<td>93.12</td>
<td>91.51</td>
<td>2.87</td>
<td>2.02</td>
<td>1.45</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>100</td>
<td>138.60</td>
<td>136.68</td>
<td>3.37</td>
<td>2.32</td>
<td>1.48</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>200</td>
<td>200.38</td>
<td>199.10</td>
<td>3.78</td>
<td>2.75</td>
<td>0.86</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>20</td>
<td>50.23</td>
<td>49.28</td>
<td>1.45</td>
<td>1.79</td>
<td>1.30</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>50</td>
<td>75.49</td>
<td>74.01</td>
<td>2.97</td>
<td>2.12</td>
<td>1.29</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>100</td>
<td>119.04</td>
<td>119.36</td>
<td>3.47</td>
<td>2.42</td>
<td>-0.24</td>
<td>16</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>200</td>
<td>183.17</td>
<td>179.26</td>
<td>3.95</td>
<td>2.82</td>
<td>2.54</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 1: Comparison studies of performance between TS and HM1

Ins.: Instance Number, Ave.: Average, OB: Objective function (1), SD: Standard deviation, Sig.: Significant
Each instance contains 10 independent tests

<table>
<thead>
<tr>
<th>Ins.</th>
<th>Number of depots</th>
<th>Number of customers</th>
<th>Ave. OB or $\overline{X}$</th>
<th>Ave. SD or $\overline{S}$</th>
<th>$T$</th>
<th>$\nu$</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>80.31</td>
<td>78.57</td>
<td>1.10</td>
<td>1.77</td>
<td>2.64</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>50</td>
<td>110.83</td>
<td>108.02</td>
<td>2.64</td>
<td>2.42</td>
<td>2.48</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>100</td>
<td>159.81</td>
<td>156.65</td>
<td>3.40</td>
<td>2.69</td>
<td>2.30</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>200</td>
<td>220.33</td>
<td>217.04</td>
<td>3.80</td>
<td>3.12</td>
<td>2.12</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>20</td>
<td>68.28</td>
<td>69.19</td>
<td>1.35</td>
<td>1.41</td>
<td>-1.47</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>50</td>
<td>93.12</td>
<td>92.64</td>
<td>2.87</td>
<td>1.66</td>
<td>0.46</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>100</td>
<td>138.60</td>
<td>134.94</td>
<td>3.37</td>
<td>3.29</td>
<td>2.46</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>200</td>
<td>200.10</td>
<td>197.38</td>
<td>3.83</td>
<td>3.72</td>
<td>1.61</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>20</td>
<td>49.73</td>
<td>49.77</td>
<td>1.45</td>
<td>1.52</td>
<td>-0.06</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>50</td>
<td>73.70</td>
<td>75.66</td>
<td>2.97</td>
<td>1.75</td>
<td>-1.80</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2: Comparison studies of performance between TS and HM2
Table 3: Comparison studies of performance between TS and HM3

<table>
<thead>
<tr>
<th>Ins.</th>
<th>Number of depots</th>
<th>Number of customers</th>
<th>Ave. OB or (X)</th>
<th>Ave. SD or (S)</th>
<th>T</th>
<th>υ</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>80.31</td>
<td>77.34</td>
<td>1.10</td>
<td>2.12</td>
<td>3.93</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>50</td>
<td>110.83</td>
<td>106.22</td>
<td>2.64</td>
<td>2.66</td>
<td>3.89</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>100</td>
<td>158.99</td>
<td>153.49</td>
<td>3.40</td>
<td>2.83</td>
<td>3.93</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>200</td>
<td>221.12</td>
<td>215.75</td>
<td>3.82</td>
<td>3.31</td>
<td>3.36</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>20</td>
<td>68.28</td>
<td>65.52</td>
<td>1.35</td>
<td>1.18</td>
<td>4.87</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>50</td>
<td>93.12</td>
<td>90.27</td>
<td>2.87</td>
<td>2.99</td>
<td>2.17</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>100</td>
<td>138.60</td>
<td>134.13</td>
<td>3.37</td>
<td>3.05</td>
<td>3.11</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>200</td>
<td>199.83</td>
<td>196.17</td>
<td>3.86</td>
<td>3.52</td>
<td>2.21</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>20</td>
<td>48.79</td>
<td>46.74</td>
<td>1.44</td>
<td>1.27</td>
<td>3.38</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>50</td>
<td>76.46</td>
<td>73.70</td>
<td>2.97</td>
<td>3.09</td>
<td>2.04</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>100</td>
<td>121.11</td>
<td>117.03</td>
<td>3.47</td>
<td>3.15</td>
<td>2.76</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>200</td>
<td>182.64</td>
<td>177.09</td>
<td>3.96</td>
<td>3.62</td>
<td>3.27</td>
<td>18</td>
</tr>
</tbody>
</table>

Now the hypothesis is tested that the population corresponding to the differences has mean $\mu$ zero. Specifically, the (null) hypothesis $\mu = 0$ is tested against the alternative $\mu > 0$. It is assumed that the OB difference is a Normal variable, and choose the significance level $\alpha = 0.05$. If the hypothesis is true, the random variable $T = (\bar{X}_1 - \bar{X}_2)/\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}$ has a $t$ distribution with:

$$\nu = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right) / \left(\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}\right)$$

degrees of freedom. The critical value of $c$ is obtained from the relation

$$\text{Prob}(T > c) = \alpha = 0.05.$$ For example, the first entry in Table 3 corresponds to the sample size $n_1 = 10^2$, $\mu_0 = 0$, sample mean for assignment algorithm and HM3 are $\bar{X}_1 = 80.31$ and $\bar{X}_2 = 77.34$ respectively. Sample standard deviation for assignment algorithm and HM3 are $S_1 = 1.10$ and $S_2 = 2.12$ respectively. Since $t = 3.94 > t = 1.77$, it is concluded that the difference is statistically significant. Table 3 displays HM3 outperforms assignment algorithm in all cases. Also in all cases differences are significant.

7. CONCLUSION

Routing and scheduling of deliveries are two crucial operational decisions in logistics distribution management. Better routing and scheduling can result in shorter delivery distance, or time, and thus, higher level of efficiency and lower delivery cost can be achieved. The VRP is used prevalently to aid the planning of these two decisions. However, the VRP is not applicable when a logistics distribution company has multiple depots. In this paper, the MDVRP was studied because the number of depots is not limited to one in many real-world situations. Besides routing and scheduling, the grouping problem is also considered in the MDVRP. Because the MDVRP integrates three hard optimization problems, three improvement techniques is introduced to challenge it. Nearest depot method, Clarke and Wright saving method and NNH for grouping, routing and scheduling of customers are used in the initial solution and improve the initial solution by three approaches. A computational study was carried out to compare HM1, HM2, HM3 and one of the strong assignment algorithm method developed by Giosa et al. Experimental results show that HM1 and HM2 outperform assignment algorithm in the most cases but all differences are not significant. HM3 completely outperforms assignment algorithm and also all differences are significant.

REFERENCES


Real-time Control of Parts Input in Flexible Production Systems by Analogy Process

Rizauddin Ramli¹*, Hidehiko Yamamoto², and Dzuraidah Abd Wahab¹

¹Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Malaysia
²Department of Human and Information Systems Engineering, University of Gifu, 1-1 Yanagido, Gifu-city, 501-1193 Gifu Prefecture, Japan

ABSTRACT

We consider a parts input problem in a flexible production system consists of 2 production lines that are an Up-Stream production line and a Down-Stream production line. The Up-Stream production line processes incoming materials by several machine tools and after processing them, delivers the parts to the Down-Stream production line via buffers. In this paper, the parts input rules to the Up-Stream production line is based on the observation of the parts line-up situation on buffers. By using the parts observation, a real-time control of parts input by pattern recognition called Real-Time Parts Input by Pattern Recognition System (RT-PIPS) is proposed. The pattern recognition consists of pattern classification process and pattern identification process. Simulations of virtual production systems have been carried out to verify that RT-PIPS is useful in real-time parts input control.

1. INTRODUCTION

One of the flexible production systems that produce multiple type parts is Flexible Transfer Lines (FTLs). Often, many FTLs operate under the Just In Time (JIT) [1-3] production policy in conditions where each of the FTLs are linked together. The conceptual JIT production policy is defined as “the basis of the Toyota production system in which the right parts are needed in the assembly line at the time they are needed, and only in the amount needed” to achieve “the absolute elimination of waste”. A JIT production system is a pull production system, in which the most well known is called the Kanban system [4]. A Down-Stream production line informs the Up-Stream production line of the needed types and amount of parts beforehand. For instance, a parent company makes a weekly pre-decided production schedule of certain parts (such as 3000 pieces of part A, 2500 pieces of part B etc.), in advance. Normally, the Up-Stream production line will schedule the part input to the production line based on the pre-decided production schedule. Pull production systems are most successful in production environments with stable demand and lead times [5, 6]. Unfortunately, when a company experiences wide variations in supply and demand from customers, the pre-decided production schedule cannot be maintained and is liable to change on the current day. In addition, an adaptive type of pull production system was proposed to monitors the actual inventory levels and adjusts the number of Kanban by releasing and capturing extra Kanban based on customer demands [7]. Moreover, Gold [8] studied a sophisticated batch service system in a push and pull manufacturing environment by using embedded Markov chain techniques.

Most of these studies only focus on a single stage system or single-product Kanban system. A mathematical method and simulation approach offline pre-decided production scheduling has also been proposed [9, 10]. Furthermore, a reactive JIT ordering system [11] for multi-stage production systems with unstable changes in demand was also proposed. However, a pre-decided production schedule is not effective if the demands from the preceding process are different from the actual schedule. In this paper, we propose a real time control of parts input control by pattern recognition called Real Time Parts Input by Pattern Recognition System (RT-PIPS). RT-PIPS observe the parts line-up situation in real-time and uses pattern recognition methods to analogize the situation. The pattern recognition method consists of two processes that are the pattern classification and the pattern identification process. In the pattern classification, the number of Out-pattern is reduced and in the pattern identification, by analogizing the situation with the pattern recognition, the most similar pattern to the parts on buffers (BUFF) can be determined. This pattern is collated with IF-THEN to acquire a rule that decides the next parts input to the bay of the Up-Stream

* Corresponding author: Tel.: (603) 8921-7022; Fax: (603) 8925-9629; E-mail: rizauddin@eng.ukm.my
Flexible Automation and Intelligent Manufacturing, FAIM2009, Teesside, UK

production line. In other words, the parts input can be decided by collating the results from pattern recognition analogical process to acquire rules for parts input. It is alike a feedback process of control system.

We adopt RT-PIPS to a virtual production system to verify that it will not negatively affect the Down-Stream production line even though the actual demands from the Down-Stream production line are different from the pre-decided schedule.

2. MODEL OF PRODUCTION SYSTEM

The production model that we consider in this paper is shown in Figure 1. This model considers a production system that consists of two types of production lines. One of the two types of the production lines is an Up-Stream production line that consists of multiple machine tools lined serially. The other is a Down-Stream production line that consists of many assembly processes. The parts are input one by one to the bay of the Up-Stream line by using the *One by One Part Input Method* [12]. The input parts are machined by its sequence by the machine tools. The parts line-up situations on the buffer (BUFF) are monitored in real-time. To realize the real-time control of parts input, a real-time control system that is based on pattern recognition method to decide the parts input to the bay of Up-Stream production line is used.

A set of Out-pattern is created randomly from the parts line-up combinations as identifiable patterns. Then, the identifiable patterns are recognized by adopting a pattern identification process. The pattern identification process is divided into two categories that are a Sum Square Method (S²M) and a Rank Replacement Method (R²M) to analogize the Out-patterns. The analogical result is then collated with parts input decisions rules to decide the next parts input rules. In the paper, we adopt IF-THEN rules as the parts input rules.

2.1. DIFFICULTIES OF DECIDING PARTS INPUT RULES

In the case where two production lines are linked and operated as a single production line, it is difficult to implement a pre-decided production schedule due to the difference between parts demands from the actual pre-decided production schedule. Solving this problem necessitates a real-time control of parts input to the bay of the Up-Stream production line during operation. However, it is quite difficult to perform a real-time control because during performing a real-time control of parts input in the flexible production system, two points of difficulties have to be settled down. The first difficulty is that the parts line-up situations on the BUFF consist of tremendous combinations, resulting in the creation of tremendous IF-THEN rules. For instance, a flexible production system that
processes \( n \) types of parts with the capacity of parts that can be lined up on the BUFF as \( b \) pieces, then it will have \((b+1)^n\) combinations.

The second difficulty relates to the constant search for accurate IF-THEN rules. From the parts input in the Up-Stream production line to the parts transferred to the BUFF, there are time lags exist between these periods. Furthermore, many parts are discharged from the production line occasionally due to machining defects, which leads to unclear situations of parts that line-up on the BUFF. Because of this reasons, the relationship between the parts situations on the BUFF and the parts input in the Up-Stream production line no longer associates as a one-to-one correspondence. It means there is no guarantee that each IF-part will correspond with one THEN-part. In the paper, both of the difficulties will be solved by our proposed RT-PIPS. The later sections of the paper describe details of the proposed RT-PIPS.

3. RULES OF ACQUISITION BY RT-PIPS

3.1 PRE-DECIDED PARTS INPUT SEQUENCE

In the paper, in order to reduce the tremendous combinations of the IF-THEN rules, we use the Out-pattern concept where smaller numbers rather than the numbers of parts line-up situations on the BUFF \((b+1)^n\) are adopted. It means that by expressing the parts line-up situations on the BUFF with the Out-pattern concept, the parts line-up situation can be limited to only \( m \) pieces of combinations, where \( m \ll (b+1)^n \). Each \( m \) pieces of the parts line-up situations are expressed by Out-pattern \( O_p(m) \).

In other words, the Out-patterns are configured as an alternative expression of the tremendous parts situations that are lined up on the BUFF. \( m \) pieces of Out-pattern \( O_p(m) \) are compared with the actual parts line-up situation as shown in Figure 2 by the pattern recognition. The pattern recognition method involves 2 processes. One is the pattern classification process. The pattern classification process is carried out of-line. The pattern classification process consist of random creation of Out-pattern set and pattern classification where the set of Out-pattern is classified based on its geometrical pattern into several Out-pattern groups. The Out-pattern groups are kept as database and is used during the analogy process of pattern identification process. The other process is the pattern identification process that is an analogy process where the Out-pattern groups are analogized on-line with the actual parts line-up situations on BUFF. We create the Out-pattern set to reduce the tremendous combinations of parts line-up situations. The Out-pattern set is then classified by the pattern classification into several Out-pattern groups.

![Figure 2 Pattern recognition of RT-PIPS](image-url)
3.2 CREATION OF OUT-PATTERN SET

In the paper, we create the Out-pattern set to reduce the tremendous combinations of parts line-up situations. TheOut-pattern set is then classified by the pattern classification into several Out-pattern groups. In the proposed RT-PIPS, in order to acquire the parts input rules by IF-THEN rules, the IF-part of the parts input rules are carried out by using the Out-pattern concept. Out-pattern is a concept of substitution of the actual parts line-up situation. In other words, the Out-pattern represents the pattern of the situation. Since the combinations of parts line-up are extremely tremendous, mathematically it is almost impossible to define the relations of the parts line-up situations and the parts input rules by mathematical formulation. So, the Out-pattern is taken as an alternative of the mathematical solutions. The creation of Out-pattern is done by generating random numbers of the product types in the FTL. The maximum BUFF size $b_{\text{max}}$ is given as a maximum number of each type $n$ of parts that can be line-up on the BUFF. By randomizing the numbers of all types of parts, the parts combinations ($PC$) on the BUFF can be determined.

$$PC = (b_{\text{max}} + 1)^n$$  \hspace{1cm} (1)

The Out-pattern for part $Op_n$ is generated randomly to the maximum BUFF size of each type of part as,

$$Op_n = \text{rand}[1,2,\ldots,b_{\text{max}}]$$  \hspace{1cm} (2)

where $n$ is the number parts types on the BUFF. The Out-pattern $O_p(m)$ is defined by the following equation,

$$O_p(m) = \{O_{p_1}, O_{p_2}, \ldots, O_{p_n}\}$$  \hspace{1cm} (3)

Then, the Out-pattern set is created by,

$$O_{\text{set}} = \{O_p(1), O_p(2), \ldots, O_p(m)\}$$  \hspace{1cm} (4)

where $m$ pieces of Out-pattern $O_p(m)$ which are far less than the actual parts combination numbers $(b+1)^n$. By using the Out-pattern, RT-PIPS describes the parts input rules as follows: the Out-pattern concept for the IF-part rules and parts input for the THEN-part rules. Furthermore, for each part that actually line-up on the BUFF, one Out-pattern corresponds to more than one possible parts input or THEN-part rule.

3.3 PATTERN CLASSIFICATION

The pattern classification process is to classify the elements of the Out-pattern set $O_{\text{set}}$ by its geometrical patterns into several specified groups as Out-pattern groups. In the paper the Out-patterns can be classify to 5 types Out-pattern groups as shown in Figure 3(a)~(e), that are the straight pattern, stair pattern, convex pattern, concave pattern and multi-concave pattern. By classifying these types of Out-pattern, the number of Out-pattern that should be analogized during the pattern identification process can be reduced far less than the tremendous numbers of parts line-up combinations. In addition, it can reduce the analogy process time of pattern identification process because only the associated Out-pattern that are classified into the specified groups will be analogized with the actual parts line-up. This means that the number of the Out-pattern $O_p(m)$ can be reduced to less than $m$ pieces. The Out-pattern group $O_{p\text{-group}}$ is expressed by the following equation.

$$O_{p\text{-group}} = \begin{bmatrix}
O_{p_{1,1}} & O_{p_{1,2}} & \cdots & \cdots \\
O_{p_{2,1}} & O_{p_{2,2}} & \cdots & \cdots \\
O_{p_{3,1}} & O_{p_{3,2}} & \cdots & \cdots \\
O_{p_{4,1}} & O_{p_{4,2}} & \cdots & \cdots \\
O_{p_{5,1}} & O_{p_{5,2}} & O_{p_{5,3}} & \cdots \\
O_{p_{6,1}} & O_{p_{6,2}} & O_{p_{6,3}} & \cdots \\
\end{bmatrix}$$  \hspace{1cm} (5)

Here, the elements of the Out-pattern groups are divided to 5 groups. Each row of the $O_{p\text{-group}}$ represents one group. The hyphens of element in the group express in which group the Out-pattern belongs to the number of the Out-pattern. Each group in the Out-pattern groups corresponds to each pattern in Figure 3(a)~5(e). A pattern classification is used before the pattern identification is done. The pattern identification process is an analogical process to compare the classified Out-pattern and the actual parts line-up situation on the BUFF. By analogizing the Out-pattern, the most similar Out-pattern can be determined.
3.4 Algorithm of RT-PIPS

In RT-PIPS, the actual parts situations that line-up on the BUFF are defined as facts that collate with the IF-part rules of the Out-pattern concept. The algorithm of RT-PIPS is done as follows.

**STEP1**: Calculate the number of parts situations Num ($P_n$) of each part $P_n$ and generate the set of parts situations numbers $S_{fact}$.

**STEP2**: Randomly create a set of Out-patterns $O_{set}$.

**STEP3**: Classify the prepared Out-pattern set by its pattern into Out-pattern groups $O_{p-group}$.

**STEP4**: Analogize between each element of $O_{p-group}$ and $S_{fact}$. Select the Out-pattern that is the most similar to the $S_{fact}$ as $O_p(M)$.

**STEP5**: Compare $O_p(M)$ and the THEN- rules and select all the parts input candidates that have the same rules. Set the candidates as $P_q$.

**STEP6**: Randomly select one part $P_{q1}$ from the overall $P_q$ as the input part.

3.5 Parts Input Rules

The THEN-part of the parts input rules describes how each Out-pattern is collated with the parts that will be input to the bay of the Up-Stream production lines. RT-PIPS selects the c pieces of the input part candidates as $I_c$ based on 5 standards as shown in Figure 4. The 5 standards indicate the priority of the input part which is judged based on the parts situations on the BUFF. Figure 4(a) ~ (e) illustrates the examples of these 5 standards respectively. In every figure, $d=d'=2$ and $P_1$~$P_{10}$ are the situations of the part line-up on the BUFF. Figure 4(a) shows a condition of [STANDARD1] where there are no parts lined up on the BUFF. In this case, each part $P_1$~$P_{10}$ can be selected as elements of $I_c$. Similarly, Figure 4(b) shows the condition of [STANDARD2]. In this case, since the difference between the desired $Opt(n)$ and the number of parts lined up on the BUFF are 2 pieces, part $P_2$ and $P_{10}$ can be selected as elements of $I_c$. Next, Figure 4(c) shows the condition of [STANDARD3] where part $P_2$, $P_3$, $P_8$ and $P_{10}$ are selected as elements of $I_c$ because these kinds of parts which are being lined up on the BUFF have not achieved the desired $Opt(n)$. Figure 4(d) shows the condition of [STANDARD4]. In this case, the difference between parts that are lined up on the BUFF and $Opt(n)$ which are less than 2 pieces can be selected as elements of $I_c$, i.e., $P_1$, $P_3$, $P_6$, $P_8$, $P_9$ and $P_{10}$. Lastly, Figure 4(e) shows the condition of [STANDARD5], where all parts can be selected as parts input candidate $I_c$. 

![Pattern Classification](image-url)
4. ANALOGY PROCESS WITH PATTERN IDENTIFICATION

The pattern identification process is to analogize the classified Out-pattern and the actual parts line-up situation on the BUFF to determine the most similar Out-pattern.

4.1 SUM SQUARE METHOD (S²M)

The smallest value of sum square of the different between the Out-pattern and the parts line-up situation on the BUFF is selected as the analogical result. The detail of the procedure is as follows:

**STEP1:** Calculate the number of the actual each parts line-up situation on the BUFF.

\[ P_i = \{p_{i1}, p_{i2}, \ldots, p_{in}\} \quad (6) \]

**STEP2:** Prepare a set of Out-pattern to be analogized with the parts line-up situation by generates randomly from the total number of combinations from \( P_i \).

\[ \text{Out}_{si} = \{O_{s1}, O_{s2}, \ldots, O_{sn}\} \quad (7) \]

**STEP3:** Subtract between the number of actual parts line-up and the Out-pattern.

**STEP4:** By using the following equation, the sum square of each parts can be calculated.

\[ X = \sum_{i=1}^{n} \left( p_i - O_{si} \right)^2 \quad (8) \]

**STEP5:** Select the smallest number of the calculated S²M as the most similar Out-pattern.
4.2 RANK REPLACEMENT METHOD (R²M)

The number of each type of the actual parts line-up on BUFF and the Out-pattern is ranked according to the smallest rank by a sorting and assimilation process of Out-pattern rank. The sorting process is as follows.

**STEP1:** List all the types of parts line-up on BUFF.

**STEP2:** Sort each type of the parts line-up on BUFF following the rank of the less number.

**STEP3:** Replace the rank of parts in Sorting-BUFF and number them as Sorting BUFF.

**STEP4:** List all the types of parts of Out-pattern and sort each type of the Out-pattern following the rank of the less number.

**STEP5:** Replace the rank of parts in Out-pattern and number them as Sorting Out-pattern.

The assimilation process is to re-rank the ranking of Sorting Out-pattern as follows:

**STEP1:** Assimilate between the new replaced S-BUFF and the new replaced Sorting Out-pattern.

**STEP2:** Change the rank of parts in Sorting Out-pattern to the same rank of the parts in Sorting-BUFF. Each time the rank is changed, one will be counted.

**STEP3:** The less counted Sorting Out-pattern is decided as the most similar Out-pattern to the parts line-up.

5. SIMULATIONS

We constructed a virtual production system and ran a simulation program. The developed RT-PIPS is applied once every 15 minutes to the virtual production system. The virtual production system consisted of a flexible production system that produced 10 types of different parts P₁, P₂, ..., P₁₀. The production ratios of the pre-decided production schedule of each part were set as, P₁ : P₂ : ... : P₁₀ = 3 : 6 : 4 : 3 : 5 : 2 : 3 : 3 : 4 : 6, respectively. In the virtual production system, the production periods were 40 hours. In order to generate a production system that was different from the pre-decided production schedule, the parts input orders were changed as the following condition.

<CONDITION-1> The parts pick-up by the Down-Stream production line is not changed from the pre-decided schedule.

<CONDITION-2> The parts pick-up by the Down-Stream production line is not so different from the pre-decided schedule.

<CONDITION-3> The parts pick-up by the Down-Stream production line is largely different from the pre-decided schedule.

These parts pick-up conditions are applied to check the effectiveness of the proposed method. The productivity of parts that are input to the Up-Stream production line by S²M and R²M are shown as Table 1 and Table 2. RT-PIPS demonstrates that although the parts pick up rules were changed; the decision of parts input can be done effectively by S²M rather than R²M.

6. CONCLUSIONS

We described the development of real-time control of parts input based on the JIT production policy by pattern recognition method. RT-PIPS can maintain stable production yields even though the production demands from the Down-Stream production line are different from the pre-decided production schedule. This means RT-PIPS that uses the Out-pattern concept collated to the IF-THEN rules is efficiently implemented.

REFERENCES


Table 1 Productivity of parts input by $S^2M$

<table>
<thead>
<tr>
<th>No</th>
<th>CONDITION-1</th>
<th>CONDITION-2</th>
<th>CONDITION-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9409375</td>
<td>0.802395833</td>
<td>0.681041667</td>
</tr>
<tr>
<td>2</td>
<td>0.936458333</td>
<td>0.801875</td>
<td>0.675729167</td>
</tr>
<tr>
<td>3</td>
<td>0.935416667</td>
<td>0.809375</td>
<td>0.678125</td>
</tr>
<tr>
<td>4</td>
<td>0.943125</td>
<td>0.7965625</td>
<td>0.677916667</td>
</tr>
<tr>
<td>5</td>
<td>0.939791667</td>
<td>0.801041667</td>
<td>0.673958333</td>
</tr>
<tr>
<td>6</td>
<td>0.936666667</td>
<td>0.809375</td>
<td>0.67375</td>
</tr>
<tr>
<td>7</td>
<td>0.932916667</td>
<td>0.808125</td>
<td>0.684375</td>
</tr>
<tr>
<td>8</td>
<td>0.937916667</td>
<td>0.803125</td>
<td>0.6771875</td>
</tr>
<tr>
<td>9</td>
<td>0.9446875</td>
<td>0.802291667</td>
<td>0.675729167</td>
</tr>
<tr>
<td>10</td>
<td>0.938333333</td>
<td>0.809208333</td>
<td>0.68125</td>
</tr>
</tbody>
</table>

Table 2 Productivity of parts input by $R^2M$

<table>
<thead>
<tr>
<th>No</th>
<th>CONDITION-1</th>
<th>CONDITION-2</th>
<th>CONDITION-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86875</td>
<td>0.685208333</td>
<td>0.5209375</td>
</tr>
<tr>
<td>2</td>
<td>0.862708333</td>
<td>0.6846875</td>
<td>0.519270833</td>
</tr>
<tr>
<td>3</td>
<td>0.874479167</td>
<td>0.695104167</td>
<td>0.512916667</td>
</tr>
<tr>
<td>4</td>
<td>0.8634375</td>
<td>0.686770833</td>
<td>0.517604167</td>
</tr>
<tr>
<td>5</td>
<td>0.859375</td>
<td>0.697916667</td>
<td>0.511875</td>
</tr>
<tr>
<td>6</td>
<td>0.870416667</td>
<td>0.690104167</td>
<td>0.512291667</td>
</tr>
<tr>
<td>7</td>
<td>0.869479167</td>
<td>0.700416667</td>
<td>0.523958333</td>
</tr>
<tr>
<td>8</td>
<td>0.87</td>
<td>0.696145833</td>
<td>0.517604167</td>
</tr>
<tr>
<td>9</td>
<td>0.8746875</td>
<td>0.694375</td>
<td>0.520104167</td>
</tr>
<tr>
<td>10</td>
<td>0.864583333</td>
<td>0.705104167</td>
<td>0.521458333</td>
</tr>
</tbody>
</table>
Quick scan of the production system for disturbances

Paulina Golinska*, Marek Fertsch, Pawel Pawlewski

Faculty of Computing and Management
Poznan University of Technology
Poznan, 60-965, Poland

ABSTRACT

In European automotive industry very often a number of different not compatible information systems is used within one company. Due to a big number of information coming from production system to production database regarding production flow, the people responsible for production planning and control are overloaded with information. The production database is not appropriate analyzed and disturbances are not identified. An approach for production system observation is needed in order to identify when disturbances appear. The aim of the paper is to propose an approach for quick scan of the production system in order to identify disturbances in production flow.

1. European Automotive Producer-Main Characteristics

Production system typical for European automotive industry is a mixture of the traditional mass production features in the departments responsible for pre-treatment (D1,D2) and lean [1,2] concept in the area of the final assembly (D3). The production process can be divided into three technological phases: body welding, painting, final assembly. The simplified schema of production system is presented at Figure 1.

The production system is characterized by high stability understand as [3]:
- steady production plans
- Master Production Schedule (MPS) is prepared many weeks in advanced
- high similarity of production routines
- high similarity of product structure (BOM), however they may differ in number of specific attributes like for example color.

The manufacturer in order to protect the stability of production system and to achieve the gains of the economy of scale combine two policies by production planning, as followed: build-to-order and build-to-forecast.

* Corresponding author: Tel.: (0048) 6187-50446; Fax: (0048) 61665-3375; E-mail: paulina.golinska@put.poznan.pl
As researches conducted in the framework of the program “3DayCar” [4] had showed in European automotive industry forecast of customer demand regarding volume and item specification is prepared many months in advance in order to figure out an attractive product mix for assembly plant balancing. Actual customer orders that are received are either fitted into the plan laid out by the production programmed months ahead, or the forecast orders in the system are amended to customer requirements. In European automotive industry production scheduling is prepared 2-3 weeks before production starts. Order scheduling, which are picked from the order bank and assigns to production weeks at the different plants. The scheduled orders for a build week are re-shuffled into a sequence of build orders for the assembly plants. In any case, only after the orders are sequenced, do suppliers actually receive their final call-off of what is required, as only then it is actually defined what parts will be needed [4]. Sequenced orders are sent to the body shop (welding department). In case most of the manufacturers the paint shop is the bottleneck of the production process so, after the body shop in order to achieve efficiency in the paint shop bodies are accumulated to be sprayed in the same color. The initial production sequence is distorted and becomes unpredictable for all subsequent operations. After paint, the cars are generally re-shuffled again before they are sent on the assembly department, to ensure the mix of cars needed for the line balancing activities. The line balancing approach leads to need for a precise control of sequencing and quick response in situation when real production sequence is disturbed and differs from assumed plans. At the same time in case of automotive industry where Just-in-Time deliveries are common business practice changes that appear in Master Production Schedule results in problems with on-time customer fulfillment. Changes in Master Production Schedules (MPS) result in due-date changes in open orders, quantity and timing for planned order of end products. Mentioned above changes are being translated into gross requirements changes for components and timing of their delivery and it effects in situation that materials needed for a particular order may not be available particularly when there is an increased in ordered quantity and shorten due date.

Number of researches that has been conducted in automotive industry regarding the improvement of the performance concentrate mainly on material management within supply chain and integration of planning activities [ISOMA 4,5,6,7]. There is a research gap for evaluation of the influence that the disturbances have on production system. The following research project has show that there is a need to elaborate a set of simply method of production system monitoring in order to easily evaluate the sequencing activities and identify potential serious disturbances that negative affects the ability to fulfill customer’s orders on time.

2. DISTURBANCES IN PRODUCTION SYSTEM

2.1. SOURCES OF DISTURBANCES

The manufacturing system is treated as a complex entity combining technological-, social-, economics subsystems and being influenced by business environment [7]. The potential sources of uncertainty in manufacturing system presents Figure 2.

![Figure 2: Sources of uncertainty in manufacturing system, modified from [8]](image)

Inputs to manufacturing system include materials, energy, and information. The inputs are being transformed into products for customer. The value adding processes in manufacturing system creates incomes for company. The
macroeconomic climate may cause the shifts in business environment which influence inputs. Manufacturing system is dynamic and contains of materials, machines and tooling, people. The outputs of manufacturing system are mainly finished goods and semi-finished goods which are being offered to customer. Many of manufacturing companies base their production plans both on due-date customer orders and demand forecast. Shifts in macroeconomics conditions might influence the veracity of forecast and may lead to serious company financial problem, freezing resources in stocks that are not required by clients.

The uncertainty conditions combined with complexity of production system increase the risk of disturbance appearance. Disturbance can be defined as an unexpected and unplanned event, which causes the deviation between planned Master Production Schedule MPS and real production flow (production sequence) within production system. The deviation between initial planned and real production sequence resulting from intentional planner activities (e.g. change of initial order priority) won’t be treated as a disturbance. Disturbances in production system can differ a lot: e.g. breakdown of machines, serious delay in deliveries, workers absence. The disturbances can be classified based on two criteria:

1. source of disturbance
2. the effect it has on production system.

Taking in consideration the first criteria the following types of disturbances, as presented on Figure 3 can be defined.

Based on the second criteria the following types of disturbances can be identified [10]:

1. production control goal disturbances
2. production capacity disturbances
3. input disturbances
4. information flow disturbances.

The main goal of production control is production process execution according to the planned MPS. The disturbances of production control function goal are any deviations in the initial MPS execution appearing at independent demand level caused by changes in customer demand or customer’s order due dates. The production resources disturbances are any variations in initial MPS execution caused by limited availability of production capacities (machines, tooling, workforce). The input disturbances are any variations in initial MPS execution appearing at dependent demand level caused by shortages of raw materials. The input disturbances are any variations in initial MPS execution appearing at dependent demand level caused by inappropriate information and data on technology: e.g. not updated Bill of Material, invalid data on maintenances, lead times, inventory levels.
Table 1: Level of particular disturbances types appearance

<table>
<thead>
<tr>
<th>Disturbances Type</th>
<th>Independent demand (MPS level)</th>
<th>Dependent demand (MRP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production control goal disturbances</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Production capacity disturbances</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Input disturbances</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Information flow disturbances</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Due to the fact that production plans (and following them MPS) in European automotive industry are prepared many weeks in advance, it can be assumed that the demand is deterministic, so only disturbances appearing at MRP level will be taken into consideration in the quick scan approach.

3. Quick Scan Approach

3.1 Approach Presentation

The main value adding process is generated by final assembly (D3). All supplies of components are supposed to arrive at final assembly department in the required sequence due to the fact that there are usually no buffers. All deliveries of raw materials and components for final assembly are based on just-in-time basis, so any changes in order sequence results in problems with on-time deliveries of components, what seriously disturbs the MPS execution. Due complexity of the production systems in the automotive industry, there are problems to identify the sources of disturbances. On the basis of the case study and the literature review, the following approach for quick scan based on few simple indicators is proposed.

The production system in automotive industry is monitored usually by number of scanners that register the data from bar codes/GS1 codes plugged to any car body at the beginning of production process. The places where scanners are located can be called production control points (PCP). Number of PCPs differs among particular manufactures. The data from PCPs is sent in real time to production databases. The structure of data depends on particular company. It can be assumed that the minimal requirements for data needed to apply quick scan approach are (Figure 5):

- Production Order Identification Number (POIN)
- PCP identification code
- Date
- Time when particular POIN passes through particular PCP

Table 2: Simplified structure of database

<table>
<thead>
<tr>
<th>Production sequence /PCP</th>
<th>PCP 1</th>
<th>PCP 2</th>
<th>PCP …</th>
<th>PCP N</th>
</tr>
</thead>
<tbody>
<tr>
<td>POIN 1</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
</tr>
<tr>
<td>POIN 2</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
</tr>
<tr>
<td>…</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
</tr>
<tr>
<td>POIN N</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
<td>Date, Time</td>
</tr>
</tbody>
</table>

The quick scan approach is based on following indicators:

- Production flow cycle indicator $XY_{Long}$ is counted as a sum of time windows between moments when analyzed Production Order Identification Number (POIN) passes through following production control points PCPs (equation 1). The time when POIN enters the production system at PCP1 is counted as 0 ($t_{pcp1} = 0$):
Quick scan of production system for disturbances

\[ \text{XY Long} = \sum_{i=1}^{n} t_{pcpi} \]  \hspace{1cm} (1)

- For block of \( k \) number of POINs the indicator is counted as (equation 2):

\[ \text{XYLong}_{\text{set}} = \frac{1}{k} \sum_{k}^{n} \sum_{i=1}^{n} t_{pcpi} \]  \hspace{1cm} (2)

In order to define the typical and critical values of production flow cycle indicator for a block of POINs on basis of long time observation of non-disturbed POINs (with the same key features as analyzed block) the “ideal” values are counted \( \text{XYLong}_{\text{setID}} \).

- Sequence dispersion range (SDR)- measured for block of POIN as a difference between maximum order identification number in defined block and minimum identification number observed at chosen production control point (equation 3).

\[ \text{SDR} = (\text{POIN max} - \text{POIN min}) \]  \hspace{1cm} (3)

The values of SDR indicators shows how defined at MPS block of POINs has been distorted by appearance of number of POINs belonging to the other blocks of orders.

- The overall production sequence quality presents the number \( k \) of POIN in analyzed block to SDR at observed production control point for example if there was “\( k \)” POIN and the SDR equals to “\( k \)” the quality of operations management regarding sequencing will be 100%. The value of this indicator is counted as following (equation 4)

\[ \text{OPSQ} = \frac{k}{\text{SDR}} \times 100\% \]  \hspace{1cm} (4)

The low value of OPQS indicator informs about serious disturbance in production flow for analyzed block. On the basis of proposed indicators it is possible to monitor production flow and to identify when serious disturbances appear and regulatory activities are needed.

2.2. THE METHOD

The quick scan approach follows mentioned below steps:
1. Define the block of orders
2. Count the OPSQ indicator value
3. Evaluate the critical value of OPSQ indicator
4. If the value of OPSQ for defined block falls below critical value then analyze the \( \text{XYLong}_{\text{set}} \) at this PCP
5. Identify the type of disturbances

<table>
<thead>
<tr>
<th>Effects caused in production system by uncertainty</th>
<th>Fall of OPSQ indicator under critical value</th>
<th>Growth of XY Long indicator over critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production capacities disturbances</td>
<td>Machines’ breakdown at the separate production process part</td>
<td>Breakdown of information systems, breakdowns of electricity supplies,</td>
</tr>
</tbody>
</table>
moters’ breakdown at the joint part of the production process

**Input disturbances**
- Non standard material requirements
- Non-standard supply lead times, poor quality, lack of needed components

**Information flow disturbances**
- Invalid production data (mainly invalid BOM)
- Changes in design of product, Non-standard design

### 3.2 Numerical Example

The numerical example presents the application of proposed quick scan approach. For the defined block of 15 orders the OPSQ indicator is calculated (Figure 6). The PCPs taken in consideration are:

- MPS-system entry the sequence is equal to initial MPS
- B1, B2- body shop (1) entry and (2) exit
- P1, P2- paint shop (1) entry and (2) exit
- A1, A2- final assembly (1) entry and (2) exit
- FC- final control

<table>
<thead>
<tr>
<th>MPS</th>
<th>B1</th>
<th>B2</th>
<th>P1</th>
<th>P2</th>
<th>A1</th>
<th>A2</th>
<th>FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPSQ</td>
<td>100%</td>
<td>27,78%</td>
<td>20,21%</td>
<td>20,83%</td>
<td>4,53%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5: OPSQ indicator average values and critical values for analyzed production system

<table>
<thead>
<tr>
<th>OPSQ av.</th>
<th>MPS</th>
<th>W1</th>
<th>W2</th>
<th>P1</th>
<th>P2</th>
<th>A1</th>
<th>A2</th>
<th>FC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100,00%</td>
<td>63,93%</td>
<td>42,05%</td>
<td>37,04%</td>
<td>46,35%</td>
<td>47,47%</td>
<td>47,95%</td>
<td>60,36%</td>
</tr>
<tr>
<td>Tolerance 25%</td>
<td>0,00%</td>
<td>15,98%</td>
<td>10,51%</td>
<td>9,26%</td>
<td>11,59%</td>
<td>11,87%</td>
<td>11,97%</td>
<td>15,09%</td>
</tr>
<tr>
<td>critical value</td>
<td>100,00%</td>
<td>47,95%</td>
<td>31,53%</td>
<td>27,78%</td>
<td>34,76%</td>
<td>35,60%</td>
<td>35,91%</td>
<td>45,27%</td>
</tr>
</tbody>
</table>

The value of OPSQ indicator is equal 100% when production flow is not disturbed. For analyzed block (Table 4) it can be observed that the flow of production is seriously disturbed (below critical value Table 5) at body shop and paint shop. At final assembly department the production sequence (production flow) finally is undisturbed but at the end of production process again the sequence is extremely disturbed (below 5% of initial sequence in MPS is kept).

Following the described in Section 3.1 procedure if the value of OPSQ for defined block falls below critical value then the analysis of indicator the XYLong_set_av at this PCP is needed.
As presented at Figure 4 the low values of OPSQ at PCP from B1 to P1 are not followed by increase of XYLong indicator over critical value. On based of Table 3 it can be assumed that production sequence is disturbed by:

1.a machines’ breakdown at the separate production process part  
2.a non standard material requirements  
3.a non-standard supply lead times,  
4.a poor quality,  
5.a lack of needed components  
6.a invalid BOM .  

The disturbances caused by non-standard production can be eliminated (2a-3a) due to the block size. According to analyzed company organizational rules non-standard production block of orders cannot be bigger than 6 POINs. In case of poor quality (4a)at this production phases the orders have to be re-process so production flow indicator would increase. Invalid BOM (6a) disturbs the production at final assembly phase not at body and paint shop. The only possible causes of deviations between planned sequence are: machines’ breakdown at the separate production process part and lack of needed components. In order to fully identified the cause a quick interview with production shift management of particular department is needed.

At production control point P2 value OPSQ indicator falls below 5%. In most of European automotive companies at paint shop orders are reshuffle in order to build up particular color block. For orders no. 8,9,0,14 to production cycle flow indicators grows over critical value. On basis of Table 3 it can be identified that disturbances in production sequence are caused potentially by:

1.b non-standard supply lead times,  
2.b poor quality,  
3.b lack of needed components  

As mentioned before the non-standard production can be excluded. At paint shop all chemicals used are rather standard and repetitive so if a lack of paint or other chemicals appear more POINs should be delayed (XYLong) indicator over critical value). It can be identified that disturbances in production sequence between P1 and P2 (at
paint shop) are caused by poor quality. Paint shop is usually a bottleneck in number of automotive companies. The characteristics of paint process results in a big amount of failures that need to be reprocess e.g. not appropriate paint coat, workers hairs or fibers from clothing stick to lack. At final assembly none disturbances have been identified. At production control point FC OPSQ value falls again below critical value. It is a place in production system when technological process is finished and car is ready to be tested. The only disturbance among lb-3b that can cause deviation in production in initial production sequence at final control is poor quality due to the fact that there are no components used. Cars (POIN 11 &15) need to be reworked, time consuming re-processing is illustrated by increased in XYLong at A2-FC part of production system.

5. CONCLUSIONS

The automotive production systems are very complex and very often it is difficult to under pressure of time to analyzed production database in order to identified disturbance. The paper presents quick scan approach that can be applied in European automotive industry. The proposed approach is designed as a help for production controllers in order to easily exclude some of potential types of disturbances by observation of few simply indicators. The application of method required the knowledge about the production system organizational conditions and main design characteristics. The further research enhances the implementation of proposed indicators into production information management system for continuous (on-line) monitoring of production plans execution. The presented research was strongly case-based oriented however due the fact that manufacturing systems of the European automotive manufacturers are similar the proposed solution might be applied as well for other companies in the sector.

REFERENCES

Application of Reverse Logistics for Remanufacturing in Electro-Mechanical Industrial Sector

Sahar El Barky, William Ion, Aziz El Sayed (*)
Design, Manufacturing Engineering Management Department
University of Strathclyde
Glasgow, G1 1XJ, UK

ABSTRACT

Recovery strategy is one the practical methods which supports sustainable objectives. Environmental policies and legislation, economic benefits and increasing customer awareness of green products are the main drivers for the extension of recovery strategies in the industrial and service sectors. Remanufacturing is preferable to recycling due to the saving of the initial value added to the raw material in producing the final product. Remanufacturing is "The process of returning a used product to at least OEM original performance specification from the customers’ perspective and giving the resultant product a warranty that is at least equal to that of a newly manufactured equivalent". The steady supply of core product is critical for remanufacturing as without it, it clearly cannot proceed. The complexity of the remanufacturing process lies in the uncertainty in the timing, quality and quantity of used products. The movement of used products from the consumer to the producer in distribution channel is defined as reverse logistics. Through reverse logistics activities, the assessment of used products is important to the overall profitability of the remanufacturing process. The decision for accepting or rejecting the used product is based on suitability for remanufacturing and is made by the dealers or OEM. Due to the lack of assessment criteria and guidelines for the assessment process, the used product assessment may be erroneous and it depends almost entirely on the experience. This paper will describe the previous work in the reverse logistics for remanufacturing. Also, it will present a proposed model showing the interaction between remanufacturing processes and reverse logistics activities. Finally, it will propose the research objectives and methodology which focus on the assessment criteria of core products and the relation between the nominal quality and acquisition price of used product. It present the initial findings of conducting industrial case study in Egypt.

1.INTRODUCTION

“Remanufacturing is the process of returning a used product to at least OEM original performance specification from the customers’ perspective and giving the resultant product a warranty that is at least equal to that of a newly manufactured equivalent” [41]. Many authors pointed out that the main drivers to the growth of remanufacturing as organized industrial sectors in many countries are as follows [7, 41, 35, and 3]:
− Legislation or take-back obligation issued by the European Union to push product recovery such as: WEEE, RoSH directive and EuP directive on the eco-design of energy-using products.
− Economic benefits by saving material and, energy cost.
− Market demand for cheaper products
− Increasing customers awareness for ‘green’ products and companies

Since the end of the 1970s, many authors have highlighted the ecological, social, and economic benefits of remanufacturing. Because of benefits, remanufacturing is considered a successful strategy for developing sustainable manufacturing. Past work gave examples of the best practice in remanufacturing strategy in the electro-mechanical sector [27, 29].

The process of remanufacturing became clarified when Ijomah developed a generic model of the remanufacturing business model, illustrating the activities of remanufacturing and their inter-relationships [40]. The activities are receiving core (used product), clean and strip the core, investigate the core, component
remanufacturing and inspection and testing. The author declared that the supply of the core (the used product), is critical for starting remanufacturing because it cannot proceed without used product to remanufacture. Also, if the supply of used products is inadequate then remanufacture, when it is able to proceed, must rely on new components rather than components cannibalized from used products [40]. Consequently, designing effective and efficient reverse supply chain system is a prerequisite for remanufacturing and a key driver for providing the economic benefits necessary to initiate and sustain customer relationship and customer loyalty [37, 4].

2. Reverse supply chain/Reverse Logistics activities

Pohlen and Farris define Reverse Logistics, guided by marketing principles, as being: “The movement of goods from a consumer towards a producer in a channel of distribution” [19]. Rogers and Tibben-Lembke describe Reverse Logistics as “the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in – process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal” [34]. Other authors state that reverse supply chain processes can be organized sequentially in five steps: product acquisition, transportation and warehousing, inspection, recovery process, and distribution and sales [23, 4, 37, and 26].

Starting with the acquisition process, Guide and Wassenhove state that product acquisition management is a key input to assessing the potential economic attractiveness of remanufacture activities [14]. This process is responsible for core acquisition in order to ensure adequate supply of cores for remanufacturing. As well, the authors rejected the idea that the firm should accept all used product returns from waste stream and show that a system for acceptance control exists. There are three major sources of used products [4]: Forward supply chain such as returns from defective or damage products and product recalls, Market driven systems, and Waste stream.

In a market driven systems, the used products returns to OEMs, retailers, dealers and third parties from the customer using different financial incentives policies such as, deposit based, credit based, buy back. Through those incentives companies are willing to obtain higher quality products for a fee. Consequently, there is less variability in the quality of used product. Large numbers of firms acquires cores directly from the customer (81.8%) by applying a trade-in or credit strategy when a remanufactured product is purchased. Those firms noticed that customer were motivated to return the products by themselves [4, 14]. Ostlin has defined other incentives such as ownership-based, service contract, direct order and voluntary based [12].

The cores are assessed in order to ensure their suitability for remanufacturing and if credit incentive is applied the price regarding to the quality of the cores are determined. Most likely, this assessment is done by the retailer or dealers. Ijomah clarified that the examination of quality of the components that make up the core are very important. Because inadequate components assessment have significant negative financial repercussions such as the discarding of good components and the use of inappropriate labour. Although, core assessment is critically important, there are a few guidelines to assist accurate component evaluation leading to excessive dependence on the experience of labours [40, 41]. The author identified the problems resulting from the shortage of documented assessment criteria and procedure as follows: inconsistency of parts inspections, excessive waste, inadequate training and poor performance monitoring. Moreover, Guide and Wassenhove claimed that the relation between nominal quality levels and acquisition price of used product should be identified in order to facilitate the effective management of core acquisition [14].
The model in figure 2 shows the interaction between reverse logistics activities and the remanufacturing process starting and demonstrates the decision points in the whole process. The model has been developed by the author from the literature and is based on two existing models: the framework for reverse logistics and a generic remanufacturing process flowchart [34, 40].

The first model of Mitra which similar to the frameworks presented by Fleischmann and Bloemhof-Ruwaard [23, 13] was selected because the model was developed upon survey on considerable number of case studies in different industries which address the design of logistics networks in a product recovery context. The model identified the recurrent reverse logistics in product recovery networks. The second model is selected because it obviously demonstrates the remanufacturing processes in details. The model validated by over 20 remanufacturing companies. Therefore, the model became a generic model for remanufacturing. The proposed model will be developed through conducting in-depth investigation of the case study.

3. PREVIOUS STUDIES OF THE REVERSE LOGISTICS

The research done in reverse logistics (RL) is mostly in practitioner-related rather than academic journals. Fleischmann notes that the literature is divided into three research areas: distribution planning, inventory control, and production planning [24]. In contrast, Dowlatshahi classified the literature according to five categories: global concepts; quantitative models; distribution, warehousing and transportation; company profiles; and applications [34]. The author stated that “the majority of the articles show lack of depth, do not describe the basic structure of Reverse Logistics, and do not define the basic concepts and terms”. Then, attention to the literature and the strategic and operational factors in Reverse Logistics systems is listed by the author. Most of scholars addressed the reverse logistics for recycling and a few for remanufacturing [15]. The following literature investigates only previous studies of Reverse Logistics network for remanufacturing.

3.1. QUANTITATIVE ANALYSIS

Simulation is a common method that is applied in quantitative methods. The method of system simulation rapidly and conveniently assesses various comprehensive reverse logistics networks under different scenarios. Consequently, an optimized network structure may be decided. Quite a few mathematical models are proposed in various literatures to optimize reverse logistics system. Fleischmann reviewed the previous quantitative models for reverse logistics network [24]. Jayaraman [38] established a mixed integer programming model that simultaneously solves for the location of remanufacturing/distribution facilities, the trans-shipment, production, and stocking of optimal quantities of remanufactured products and cores. The authors discussed the managerial uses of the model for logistics decision-making.

Krumwiede and Sheu developed a reverse logistics decision-making model to guide the process of examining the feasibility of implementing reverse logistics in third-party providers such as transportation companies [5]. Lourenço and Pablo developed a medium term production planning model dealing with the two new concepts: Partnerships and Reverse Logistics [10]. Jayaraman proposed a mathematical programming model for the management of product return flows, induced by various forms of reuse of products and material [37]. Pochampally proposed a newly developed method called physical programming to identify potential facilities in a set of candidate recovery facilities operating in a region where a reverse supply chain is to be established [17].

Savaskan addressed the problem of choosing the appropriate reverse channel structure for the collection of used products from customers for remanufacturing [31]. The authors modeled the three options for collecting the used product namely from the manufacturer, the retailer and third party. The authors found that the most effective method of product collection activity for the manufacturer is the retailer. Beamon cited by Mutha and Pokharel have developed an integer programming model for a four level reverse supply chain by assuming infinite storage capacities and same holding costs for recovered and new products [2,1]. In the same year, Kusumastuti proposed a multi objective and multi-period mixed integer linear programming model for network design for modularized products. The authors used the model to determine the number of existing forward flow facilities and the number of dedicated facilities to be set up for handling return flow [28]. Kim have proposed a general framework for remanufacturing environment and a mathematical model to maximize the total cost savings by optimally deciding the quantity of parts to be processed at each remanufacturing facilities, the number of purchased parts from subcontractor [16].
Figure 2: Proposed model demonstrates the interaction between reverse logistics activities and remanufacturing based on Mirta [34] and Ijomah [40].
Kara has presented a simulation model of a reverse logistics networks for collecting EOL electrical appliances in the Sydney Metropolitan Area [18]. The simulation results showed that the model calculates the collection cost in a predictable manner. Moreover, it provides a tool to understand how the system behaves by carrying out what-if assessments. Finally, the authors suggested that the low cost can be achieved when local councils act as collectors. In the same year, Listes proposed a generic stochastic model for the design of networks comprising both supply and return channels, organized in a closed loop system [26].

Lu and Bostal proposed mixed integer model to solve a two-level location problem with three types of facility to be located in a specific reverse logistics system, namely a ReManufacturing Network (RMN) [42]. The model considers simultaneously the forward and reverse flows and their common interactions. Wojanowski presented a continuous modeling framework for designing a drop-off facility network and determining the sales price that maximize the firm’s profit under a given deposit–refund [30]. Lee and Dong have developed a deterministic programming model for systematically managing forward and reverse logistics flows [30]. Srivastava provided an integrated holistic conceptual framework that combines descriptive modeling with optimization techniques at the methodological level [36].

**QUALITATIVE ANALYSIS AND CASE STUDIES**

De Brito reported that there are more 60 case studies for reverse logistics for different recovery processes. Approximately 60% of the cases are in the manufacturing category; about 20% are within wholesale and retail trade and about 10% in construction. However, to a certain extent a few authors investigated reverse logistics network for reverse logistics for remanufacturing field using case studies. It is observed that almost half of them deal with metal products, machinery and equipment and the majority of the cases are on products with high value [21].

Thierry declared that Recovery options that need take back of product may be classified as two kinds: direct recovery and process recovery. Recovery options can be summarized from various publications as: reuse, repair, remanufacturing, recycling and disposal [25].

Dijkhuizen discussed the remanufacturing network of IBM. The author dealt with the problem of where to re-process the products: in each country, or centrally at one place in Europe [9]. Meijer discussed the remanufacturing of used scanners, printers, copiers, faxes at Canon [8]. Krikke analyzed reverse logistics for remanufacturing of photocopiers [11]. The authors proposed a multi integer linear programming model to determine optimal location for preparation and reassembly operations. Guide clarified that some suppliers for toner cartridges, including UNISYS, deliver their cartridge in a box that can be returned for free to them or via another third party logistics service provider like Hewlett Packard or Xerox [15]. Also, Guide and Wassenhove discussed acquisition price incentive through a U.S. cellular phone remanufacturer that is also very active in setting prices to buy used mobile phones in the Business to Business environment [14].

Toktay et al described special example of service returns on Kodak’s single use camera. The authors proposed a closed queuing network model to determine a cost-efficient order policy for the external supplies. Major difficulties are largely unknown and difficult to observe from return probabilities and market distribution [20]. They assessed the importance of information on the returns for the control of the network. Fleischmann et al answered the following question: What activities are involved in reverse logistics? The authors identified main differences between the recovery networks: links with other networks; open vs. closed loop structure; and degree of branch cooperation. Moreover, they pointed out the classification scheme for different types of recovery networks: bulk recycling network; assembly product remanufacturing network; and re-usable item network. The authors proposed generic facility location model which is used to analyze the impact of product return flows on logistics networks [22].

De Brito and Dekker answered the following question: Why do companies pursue reverse logistics? The authors categorized the driving forces under three headings: Economics, legislation and extended responsibility. Additionally, reverse logistics is the main source for supplying used products for the companies in order to establish product recovery actions such as remanufacturing [21]. Ravi et al presented an analytic network process (ANP) based decision model to analyze the options in reverse logistics (RL) for end-of-life computers and link them to the determinants, dimensions and enablers of reverse logistics (RL) [29]. The model links the financial and non-financial, tangible and intangible, internal and external factors, thus providing a holistic framework for the selection of an alternative for the reverse logistics operations for end of life (EOL) computers.
4. SUMMARY OF THE LITERATURE

The literature shows that the assessment process of used product during reverse logistics is significant research problem as the incident leading to model development could be unique. Therefore, the research emphasizes on developing an integrated model for reverse logistics to remanufacturing in electro-mechanical sector. The model identifies the optimal criteria for acceptance the suitable used product in hierarchical form.

5. RESEARCH METHODOLOGY

To find appropriate companies for being case studies, a background study carried out through collecting data via semi-structured interviews with many companies in electro mechanical sector in Egypt. These interviews were conducted via telephone and emails in order to find out which of those companies are apply reverse logistics system for remanufacturing. The case study is an authorized dealer for Caterpillar as an OEM. Caterpillar is one of the leaders in applying reverse logistics for remanufacturing strategy all over the world. The choice of Case Study Company was made based on the following: Annual remanufacturing volumes; Relation to OEMs; Product complexity and accessibility. The research follows an inductive pattern because the study is based on a transition from specific observation and practical collected data from case study to broader generalizations and ultimately theories.

In the initial investigation, the reverse logistics current system is investigated supported by observation, documentation and interviews with key company personnel. The next section shows the application of credit strategy and the criteria which used by dealers during the assessment process. The work will compare the initial findings of the current system in the case with the proposed model from the literature in order to identify the gaps between the literature and practical application. The gap analysis will be the basis for developing the proposed model. The developed mode will be validated through identifying different scenarios, and running considerable number of experiments using focus group. The research paradigm is integrated between qualitative and quantitative data in order to establish the model because it is useful to explore a host of factors that may be affecting reverse logistics activities which not addressed obviously in the literature.

6. APPLICATION OF CORE CHARGE/CREDIT PHILOSOPHY

According to Östlin, the customers can receive credits for the type of cores they supply to the company. As a result, this means that the customer can return cores without purchasing a remanufactured product. The amount of credit the customers gain from the supplied cores is dependent on the state of the core, and the credit they have acquired gives them a discount when ordering a new remanufactured product [12].

![Diagram of core and credit flow](image)

Figure 3: Illustration of the core and credit flow between customer, dealers and remanufacturer

Credits are given according to two factors: first, the quality level of the core, and if any of the specific components in the core are missing; second, the amount of credit given, is also variable between different types of products, with cores in higher demand given a higher credit, and lower credit given for cores with lower demand. In this way, the remanufacturer gets a variety of cores and can practice some level of control through the credit system, while the customer can return cores for credit. This type of system enables the remanufacturer to control the balance between the supply and demand. The credit system can also work as a method for assessing the incoming quality level of the cores according to the number of credits given for returned cores. Comparable to, The initial findings of the interviews shows that the case study company applies “core charge/credit” philosophy in order to the following reasons:

– Ensure the timely return of cores fundamental to the successful operation of an exchange program.
– Prevent unplanned discounts which would result if users purchased Remanufactured Products without intending to return an acceptable one.

– Encourage the customer to return the core and buy the remanufactured product.

Cores qualified for exchange credit are those received by the dealers from the sale of corresponding remanufactured products to users. Therefore, the quantity of exchange credits for which a dealer is eligible will not exceed the quantity of corresponding remanufactured products purchased from the company. The types of core credit which are available for dealers are classified to the following:

– Exchange credit: when the core received by the dealer resulted from the sales of a corresponding remanufactured product purchased from the company.

– Direct purchase credit: for a core which resulted from other than a sale of a remanufactured product.

– Warranty credit: for a core resulting from the use of the company remanufactured product in warranty repair.

– Dealer surplus returns: Remanufactured products are returnable to the company on the same basis as new products.

7. CORE ACCEPTANCE CRITERIA GUIDELINES IN ELECTRO-MECHANICAL SECTOR

The core acceptance criteria used by dealer during inspection process to determine if a core is eligible for exchange credit. Table 1 shows the acceptance criteria for camshafts which concluded from the interviews and the documentation provided. Also, the price is evaluated according to level of deposit refund which can be full core deposit refund or damage core refund. For instance, a few product groups, such as camshafts have two levels of possible core deposit refunds. In some circumstances, the core credit will be reduced if major parts are missing or damaged in the core. After inspection, the proper core handling, packaging and labeling are done and the cores are stored until transporting to remanufacturing center.

Table 1 the acceptance criteria and different levels of exchange credit for camshafts (regarding to CAT REMAN Guide TEPs dealer guide and interview with key personal in Mantrac)

<table>
<thead>
<tr>
<th>level of exchange credit</th>
<th>Camshafts acceptance criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Core Deposit Refund</td>
<td>– Acceptable part number (company part number only)</td>
</tr>
<tr>
<td></td>
<td>– Not visibly cracked, chipped, broken, chromed or welded</td>
</tr>
<tr>
<td></td>
<td>– Lobes not visibly chipped, missing material, or have signs of case crushing.</td>
</tr>
<tr>
<td></td>
<td>– Journal not visibly spun, chipped, or missing material.</td>
</tr>
<tr>
<td></td>
<td>– No signs of non-operational damage (mishandling, excessive rust, corrosion, pitting, or evidence of unsuccessful attempts to salvage).</td>
</tr>
<tr>
<td></td>
<td>– Fully assembled and complete</td>
</tr>
<tr>
<td>Damaged Core Refund</td>
<td>– Acceptable part number - Caterpillar part</td>
</tr>
<tr>
<td></td>
<td>– Lobes not visibly chipped, missing material, or have signs of case crushing.</td>
</tr>
<tr>
<td></td>
<td>– No signs of non-operational damage (mishandling, excessive rust, corrosion, pitting, or evidence of unsuccessful attempts to salvage).</td>
</tr>
<tr>
<td></td>
<td>– Fully assembled and complete (assemblies &amp; kits only. see add charges)</td>
</tr>
<tr>
<td>No credit</td>
<td>– Unacceptable part number or not a Caterpillar part.</td>
</tr>
<tr>
<td></td>
<td>– Non-Operational damage (mishandling, excessive rust, corrosion, pitting or evidence of unsuccessful attempts to salvage).</td>
</tr>
<tr>
<td></td>
<td>– Visibly cracked, broken, chromed, or welded.</td>
</tr>
<tr>
<td>Add Charges</td>
<td>– Add charges will be applied for any missing or damaged gear, rocker arm, or lifter</td>
</tr>
</tbody>
</table>
8. CONCLUSION

The above discussion shows that topics which need to be covered in reverse logistics not only the networking structure and inventory analysis. The collection of used products, their pricing and their suitability to recovery strategy through an establish system are key area for further research.

This paper has illustrated the previous work in the reverse logistics for remanufacturing. The proposed model demonstrated the interaction between reverse logistics activities and remanufacturing processes. Also, the paper explained the application of credit strategy in reality as incentive strategy to return back used product in electro-mechanical sector. Finally, the core acceptance criteria are presented to help dealers and OEMs during core assessment process through conducing case study. In future, an integrated model of reverse logistics for remanufacturing will be developed. The model will identify optimal criteria for acceptance the suitable used product in hierarchical form based on the practical data of different products based on conducting the case study and literature.

9. ACKNOWLEDGEMENT

The work acknowledges core management director of Mantrac Egypt Mohamed Salah El Din for his cooperation.

10. REFERENCES

A Heuristic Strategy for Stochastic Job-Shop Scheduling with Machine Breakdowns

M. Morady Gohareh\textsuperscript{1}, R. Ghasemy Yaghin\textsuperscript{1}, S.M.T. Fatemi Ghomi\textsuperscript{1*}

\textsuperscript{1}Department of Industrial Engineering
Amirkabir University of Technology
Tehran, 15916- 34311, Iran

\textbf{ABSTRACT}

This paper considers a stochastic job-shop scheduling problem. Many real world scheduling problems address probabilistic behavior in process times, due dates or other parameters. A job-shop scheduling problem was considered with stochastic process times and possible machine breakdowns. Distribution functions of process times are supposed to be known. Also, life times of machines are assumed to be exponentially distributed. In stochastic scheduling problems, unlike the deterministic variants, the solution cannot be fully determined a priori. In other words, every time a machine becomes available, a job has to be chosen to be passed on it from all the jobs that are waiting in line. The objective usually would be minimization of some measure consisting total tardiness and earliness. A solution to such a problem consists of two parts:

1. A strategy (rule) to choose one job from the jobs that are waiting in line
2. Near optimal values of start times of jobs

A heuristic rule based was devised on the famous central limit theorem to choose the job that should be passed on the freed machine. Also, an extensive simulation was used and a special push-forward technique to optimize start times and evaluate our solution method.

\section{INTRODUCTION}

Job-shop scheduling is a well-known NP-hard problem which has spurred an extensive and rapidly growing body of literature [1]. However, when uncertain parameters are included in the problem, very few works are addressed [2-7]. Uncertainty can be incorporated in Job-shop scheduling as stochastic or fuzzy parameters and on-line facets. Also the problem can be static, where a schedule is generated once, and dynamic, where the solution may or should be modified in discrete time horizons.

When parameters of Job-shop scheduling are random variables of known probability distributions, the problem belongs to the wide category of stochastic combinatorial optimization problems. For instance due dates, process times, cost parameters and technological precedence of operations can be random variables. In this case, the objective function may involve quantities such as expected value and variance of tardiness and inventory cost, probability of violation of due date constraints and so on. A solution method for stochastic job-shop scheduling is discussed in the context of stochastic optimization. Despite of being hard to solve, stochastic job-shop scheduling problem is very appealing to be challenged because most real-world situations include some levels of randomness.

As mentioned before, in stochastic combinatorial optimization problems uncertain information is described by means of random variables of known probability distributions. A more precise definition of stochastic combinatorial optimization problems is given in [8]. Static stochastic combinatorial optimization problems are characterized by the fact that decisions, or equivalently, the identification of a possibly optimal solution is done before the actual realization of random variables. This framework is applicable when a given solution may be applied with no modifications (or very small ones) once the actual realizations of random variables are known. This type of problems is also addressed in the literature as a-priori optimization. Dynamic stochastic combinatorial optimization problems arise when it is not possible or convenient to design a solution that is usable as it is for any realization of

\textsuperscript{*}Corresponding author: Tel. : (+98) 2166413034; fax: (+98) 2166413025; E-mail: fatemi@aut.ac.ir
the random parameters. In this case, decisions must be taken also after the random events have happened. This could also be done in stages, because it is often the case that the uncertain information is not revealed all at once, but in some stages. Roughly expressed, a dynamic stochastic combinatorial optimization problem is the one where decisions are taken at discrete times \( t = 1, 2, \ldots, T \), the horizon \( T \) being finite or infinite. In dynamic stochastic combinatorial optimization problems the concept of solution used in the static variants, which is similar to the concept of solution in their deterministic variants, is no longer valid. Here a solution is a policy or strategy that is adopted at each time horizon \( t \). In other words, a policy is a set of rules that says what action will be taken for each possible random future event.

In job-shop scheduling, probably the first thing that comes in to mind regarding stochastic facets, is the situation where process times are random variables. In such a problem, it is not possible to decide the precedence for processing jobs on each machine beforehand (unlike the deterministic variants). In other words, at the time that a machine is freed, a decision should be made to specify the job that should be passed on it. This calls for devising a rule or strategy (like the famous “first come first service” rule) to choose a job. While there are some works that have considered stochasticity in job-shop scheduling [2-7], the only paper was found that addresses this rule-based situation is [2]. In this paper, choosing a job from the line is carried out by using a cost-based decision-making rule. A pair wise comparison is raised among the competing jobs (jobs that are waiting in line for the freed machine) and the winner is passed on the machine. The decision-making rule, which is the basis of the competition, is carried out by obtaining the expected value of total costs of competing jobs. This paper uses the central limit theorem to estimate the costs. However, in [2] machine breakdowns are not considered. The same approach was used as [2] but it also is assumed that machines may breakdown during processing jobs.

2. DESCRIPTION OF THE PROBLEM

The problem considers a manufacturing cell with a set of \( N \) jobs and a set of \( M \) machines. Each job (order) consists of a sequence of \( M \) operations, each of which needs to be processed on a unique given machine. Preemption is not allowed but if a machine breaks down while it is processing a job, it should continue that job right after being repaired. Life times of machines are exponentially distributed. Now it is defined the following notations \((1 \leq i \leq M, 1 \leq j \leq N)\):

- \( p_{ij} \): random variable representing the process time of the \( i \)'th operation of job \( j \)
- \( \bar{p}_{ij} \): expected value of \( p_{ij} \)
- \( \sigma_{ij} \): standard deviation of \( p_{ij} \)
- \( y_i \): exponential random variable representing the life time of machine \( i \)
- \( \lambda_i \): the parameter of the exponential distribution of \( y_i \), i.e. \( f_i(y) = \lambda_i e^{-\lambda_i y} \)
- \( f_i \): the fixed value of repair time of machine \( i \)
- \( d_j \): due date of job \( j \)
- \( c_j^1 \): the penalty that should be paid for each time unit that job \( j \) is finished before \( d_j \)
- \( c_j^2 \): the penalty or tardy cost that should be paid for each time unit that job \( j \) is finished after \( d_j \)
- \( c_j^3 \): the penalty or tardy cost that should be paid for not delivering job \( j \) on time (to be paid once to the customer)
- \( St_j \): the start time of job \( j \), i.e. job \( j \) is not allowed undergoing any process sooner than \( St_j \) (decision variable)

Our objective is to obtain two issues in order to minimize the total inventory and tardy costs: a heuristic strategy or rule that specifies which job should be passed on a machine once it is freed (however if at most one job is ready to be processed on the freed machine there is nothing to decide), and near optimal values for \( St_j \)'s.

The problem was defined here is a dynamic or, as it is called in the literature, a reactive job-shop scheduling. Such a job-shop problem may be shown as “\( \sum_i w_i \sum_j T_j^i + w_j E_j^j \)” [9]. In problems that tardy and inventory costs exist, the issue of determining \( St_j \)'s arises and this is the situation we are dealing with here.
3. HEURISTIC STRATEGY

The heuristic strategy we are about to develop is similar to the decision-making rule with cost parameters method devised in [2]. However, major differences could be seen due to our approach and the fact that machines may break down. Assume that at time horizon t a set of jobs, namely \( C^j_{m} \), are competing for machine m. Now let us introduce some notations:

- \( m_{ij} \): the machine that processes the k'th operation of job j, \( 1 \leq k \leq M, 1 \leq j \leq N \)
- \( p'_{ij} \): the operational time that job j will be in the shop floor from the time its \( (x-1) \)'th operation is finished, where \( j \in C^j_{m} \) (\( m = m_{ij} \))
- \( \bar{p}'_{ij} \): expected value of \( p'_{ij} \)
- \( \sigma'_{ij} \): standard deviation of \( p'_{ij} \)
- \( C^j_{i} \): total cost of job j if at time horizon t job j is passed on machine m and on condition that job j never waits in line in future where \( j \in C^j_{m} \)
- \( C^j_{kij} \): total cost of job j if at time horizon t job \( j_k \) is passed on machine m and immediately after this event job \( j_1 \) is passed on machine m and never waits in line for any other machine in the future, \( j_1, j_2 \in C^j_{m} \)

Expected value of \( C^j_{i} \) and \( C^j_{kij} \) are calculated in Section 3.1. In the definition of \( C^j_{i} \) and \( C^j_{kij} \), it has been assumed that \( j \) and \( j_k \), respectively, will not wait in any line in future. This assumption, despite of being very critical for the following analysis and also the analysis that it will be carried out in section 3.1, is not realistic. However, as we are devising a heuristic rule (not an exact one), the result is applicable for setting up a comparison framework between jobs. Besides, one should pay attention that the assumption, although being unrealistic, is supposed to hold for all jobs that are in the same position and this creates a fairer situation. Now let us assume that jobs \( j_1, j_2, \ldots, j_n \) are competing for machine i at time horizon t. If \( E[C^1_i] + E[C^1_{k_i} < E[C^1_{k_i}] + E[C^1_{k_h}] \), then it is inferable that in a comparison between job \( j_1 \) and job \( j_2 \), it is probably less costly to select \( j_1 \). Further, the winner can be set up against \( j_3 \) and this pairwise comparison can be continued to the last job. The final winner is passed on the machine.

3.1. ON COMPUTING \( E[C^1_i] \) AND \( E[C^1_{k_h}] \)

To compute these cost parameters, like [2], we make use of the well known central limit theorem. It can be understood if now machine breakdown could happen, then \( \bar{p}'_{ij} = \sum_{i=x}^{M} \bar{p}'_{ij} \) and \( \sigma'_{ij} = \sqrt{\sum_{i=x}^{M} \sigma^2_{ij}} \). However, by conditioning on the number of breakdowns during the process of job j, one can prove the following two equations:

\[
\bar{p}'_{ij} = \sum_{i=x}^{M} \left[ \frac{1}{\lambda_{m_{ij}}} + f_{m_{ij}} \right] \left( 1 - p'^*_{ij} \right) + \bar{p}'_{ij} \tag{1}
\]

\[
\sigma'_{ij} = \sqrt{\sum_{i=x}^{M} \left[ 1 - p'^*_{ij} \right] \lambda^2_{m_{ij}} + \left( \frac{2}{\lambda_i} + f^2_{m_{ij}} + \frac{2f_{m_{ij}}}{\lambda_{m_{ij}}} \right) \left( 1 - p'^*_{ij} \right) \left( \frac{1}{\lambda_{m_{ij}}} + f_{m_{ij}} \right)^2 \left( 1 - p'^*_{ij} \right)^2 + 2 \left( \frac{1}{\lambda_{m_{ij}}} + f_{m_{ij}} \right)^2 \left( 1 - p'^*_{ij} \right)^3 + \sigma^2_{p^*} \left( 2 - p'^*_{ij} \right) + \sigma^2_{\sigma'^*} \right] \tag{2}
\]
where \( p_{ij}^* = \Pr[y_{m_i} > p_{ij}] \).

Now, according to the central limit theorem, one can say that if \( M - x + 1 \) is large enough then \( p'_{xj} \sim N(\tilde{p}_{xj}, \sigma'_{xj}) \) where \( N(\tilde{p}_{xj}, \sigma'_{xj}) \) is the normal density function with \( \tilde{p}_{xj} \) and \( \sigma'_{xj} \) as mean and standard deviation. Because \( p'_{xj} \geq 0 \), it is more appropriate to write \( p'_{xj} \sim N(\tilde{p}_{xj}, \sigma'_{xj}) (z \geq 0) = \frac{N(\tilde{p}_{xj}, \sigma'_{xj}) (z)}{1 - \Phi \left( \frac{-\tilde{p}_{xj}}{\sigma'_{xj}} \right)} \) where \( \Phi(z) \) is the cumulative distribution function of standard normal. For simplicity we use \( N(z) \) instead of \( N(\tilde{p}_{xj}, \sigma'_{xj}) \).

Now we have enough tools to compute \( E[C_i^x] \):

\[
C_i^x = \begin{cases} 
(d_j - t - p'_{xj})c_j^1 & \text{if } d_j - t - p'_{xj} \geq 0 \\
(p'_{xj} + t - d_j)c_j^3 + c_j^3 & \text{if } d_j - t - p'_{xj} < 0 
\end{cases} \implies 
\]

\[
E[C_i^x] = E[C_i^x \mid p'_{xj} \leq d_j - t] \Pr[p'_{xj} \leq d_j - t] + E[C_i^x \mid p'_{xj} > d_j - t] \Pr[p'_{xj} > d_j - t] = 
\]

\[
E[C_i^x \mid p'_{xj} \leq d_j - t] \int_{d_j - t}^{d_j - t - 1} N(z)dz + E[C_i^x \mid p'_{xj} > d_j - t] \int_{d_j - t}^{d_j - t - 1} N(z)dz \implies 
\]

After some algebraic operations and integrations, we can write:

\[
E[C_i^x] = c_i^1 \cdot d_j - t - \left( \Phi \left( \frac{d_j - t - \bar{p}_{xj}'}{\sigma'_{xj}} \right) - \Phi \left( \frac{-\bar{p}_{xj}'}{\sigma'_{xj}} \right) \right) - \left( \Phi \left( \frac{d_j - t - \bar{p}_{xj}'}{\sigma'_{xj}} \right) - \Phi \left( \frac{-\bar{p}_{xj}'}{\sigma'_{xj}} \right) \right) \times 
\]

\[
\Phi \left( \frac{d_j - t - \bar{p}_{xj}'}{\sigma'_{xj}} \right) - \Phi \left( \frac{-\bar{p}_{xj}'}{\sigma'_{xj}} \right) + 1 - \Phi \left( \frac{-\bar{p}_{xj}'}{\sigma'_{xj}} \right) 
\]

\[
\left( \left( \frac{\bar{p}_{xj}'}{\sigma'_{xj}} \right) 1 - \Phi \left( \frac{d_j - t - \bar{p}_{xj}'}{\sigma'_{xj}} \right) \right) + \frac{\sigma'_{xj}}{\sqrt{2\pi}} e^{-\frac{(d_j - t - \bar{p}_{xj})^2}{2\sigma^2_{xj}}} - c_j^3 \right) + t - d_j \right) c_j^3 + c_j^3 \right) \times \frac{1 - \Phi \left( \frac{d_j - t - \bar{p}_{xj}'}{\sigma'_{xj}} \right)}{1 - \Phi \left( \frac{-\bar{p}_{xj}'}{\sigma'_{xj}} \right)} \]

(3)

In order to calculate \( E[C_{i,kl}^x] \), the very same steps should be taken. It is easily understood that:
\[
C_{hj} = \begin{cases} 
(d_h - t - p'_{xh} - p_{oh})c_{ij}^h & \text{if } d_h - t - p'_{xh} - p_{oh} \geq 0 \\
(p'_{xh} + t + p_{oh} - d_h)c_{ij}^h + c_{ij}^3 & \text{if } d_h - t - p'_{xh} - p_{oh} < 0 
\end{cases}
\]

where \( m = m_{xh} \) and \( m = m_{oj2} \). Here we can write \( p'_{xh} + p_{oh} - N_{\bar{p}_{xh}, \bar{p}_{oj2}, \sigma_{xh}, \sigma_{oh}}(z) \) and then follow the same way we did for equation 3 to achieve an equation for \( C_{hj2} \). The desired result, which we do not bring here due to brevity, is obtained by replacing \( \bar{p}_{xh} \) by \( \bar{p}_{xh} + \bar{p}_{oj2} \) and \( \sigma_{xh} \) by \( \sigma_{xh} + \sigma_{oh} \), in equation 3.

4. Optimizing \( St_j \)'s

In order to optimize \( St_j \)'s we use a simple push forward algorithm. Roughly expressed, we consider jobs one by one (no special order), and we try to push the \( St_j \)'s forward as much as the total cost of the manufacturing cell decreases. The process of determining the costs for each \( St_j \) is done via simulation. In each step of the algorithm we increase the \( St_j \) for the considered job by a discretization step (since the \( St_j \)'s are continuous) and then use simulation to evaluate the total cost. Here, we define the following notations:

- \( DS_j \): discretization step for \( St_j \)
- \( I_{\text{max}} \): maximum number of simulation runs for each value of \( St_j \)
- \( St_j^s \): start time of job \( j \) in step \( s \) of the optimization algorithm
- \( TC_j^s \): mean total cost of shop floor in step \( s \) of the optimization algorithm while \( St_j \) is being optimized

Now, we propose the following algorithm:

1. let \( j=1 \) and \( s=1 \)
2. for \( J=1 \) to \( N \) let \( St_j^s = 0 \)
3. run the simulation model for \( I_{\text{max}} \) iterations and calculate \( TC_j^s \)
4. if \( j \leq N \) then
   \[ St_j^s \leftarrow St_j^s + DS_j, s \leftarrow s + 1 \]
   Else
   stop
5. run the simulation model for \( I_{\text{max}} \) iterations and calculate \( TC_j^s \)
6. if \( TC_j^s > TC_j^{s-1} \) then
   \[ St_j^s \leftarrow St_j^{s-1}, j \leftarrow j + 1, s \leftarrow 0, St_j^s \leftarrow -DS_j \] and go to 4
   Else
   go to 4

5. Computational Experiments

In order to verify the efficiency of the optimization algorithm and the heuristic strategy, experiments have been carried out. To have a suitable basis for comparison, the exact test problems used in [2] are also used here, that would be a job-shop manufacturing cell comprising 16 jobs and 8 machines. Since we are considering machine breakdowns, we do not expect the results to be as good as what is reported in [2]. However we bring the results of [2] to have a minimal basis of comparison. Process related parameters are given in Table 1. The data is presented in 3-tuple format \((x, y, z)\).

Table 2 contains cost parameters and due dates. Operations should be processed on machine \( x \) while three distributions of random process time have been considered:

1. Normal distribution with the mean \( y \) and the variance \( z \)
2. Uniform distribution on the interval \([y - 3\sqrt{z}, y + 3\sqrt{z}]\)

3. Exponential distribution with the mean \(y\)

Our method was implemented on an Intel® core™2 CPU T5600 that works at 1.83GHz with one gigabyte of RAM. The simulation model of the shop floor was run for 200 simulations for each test problem (\(I_{\text{max}} = 200\)).

Table 1: Data regarding process times and technological precedence of operations

<table>
<thead>
<tr>
<th>Operations</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jobs</td>
<td>1,140,100</td>
<td>4,120,100</td>
<td>1,160,100</td>
<td>8,400,400</td>
<td>6,300,225</td>
<td>5,250,225</td>
<td>2,120,100</td>
<td>7,190,100</td>
</tr>
<tr>
<td></td>
<td>4,180,100</td>
<td>1,160,100</td>
<td>8,500,400</td>
<td>6,300,225</td>
<td>3,250,100</td>
<td>2,180,100</td>
<td>7,170,100</td>
<td>5,400,225</td>
</tr>
<tr>
<td></td>
<td>2,160,100</td>
<td>5,190,225</td>
<td>1,600,400</td>
<td>4,100,100</td>
<td>3,140,100</td>
<td>7,150,100</td>
<td>8,600,400</td>
<td>6,400,225</td>
</tr>
<tr>
<td></td>
<td>1,180,100</td>
<td>2,200,100</td>
<td>4,140,100</td>
<td>8,400,400</td>
<td>5,190,225</td>
<td>6,320,225</td>
<td>7,170,100</td>
<td>3,160,100</td>
</tr>
<tr>
<td></td>
<td>5,225,225</td>
<td>4,140,100</td>
<td>1,160,100</td>
<td>8,540,400</td>
<td>6,280,225</td>
<td>3,280,100</td>
<td>2,100,100</td>
<td>7,200,100</td>
</tr>
<tr>
<td></td>
<td>1,170,100</td>
<td>2,180,100</td>
<td>8,480,400</td>
<td>5,380,225</td>
<td>3,150,100</td>
<td>4,130,100</td>
<td>7,180,100</td>
<td>6,290,225</td>
</tr>
<tr>
<td></td>
<td>5,400,225</td>
<td>4,150,100</td>
<td>1,580,400</td>
<td>8,600,400</td>
<td>7,160,100</td>
<td>6,300,225</td>
<td>3,200,100</td>
<td>2,190,100</td>
</tr>
<tr>
<td></td>
<td>7,110,100</td>
<td>8,500,400</td>
<td>1,600,400</td>
<td>2,190,100</td>
<td>5,400,225</td>
<td>4,100,100</td>
<td>6,300,225</td>
<td>3,180,100</td>
</tr>
<tr>
<td></td>
<td>1,100,100</td>
<td>2,500,400</td>
<td>3,250,100</td>
<td>5,300,225</td>
<td>6,210,225</td>
<td>4,120,100</td>
<td>8,650,400</td>
<td>7,090,100</td>
</tr>
<tr>
<td></td>
<td>3,100,100</td>
<td>8,400,400</td>
<td>4,100,100</td>
<td>1,400,400</td>
<td>2,100,180</td>
<td>5,220,225</td>
<td>6,310,225</td>
<td>7,130,100</td>
</tr>
<tr>
<td></td>
<td>5,240,225</td>
<td>7,140,225</td>
<td>8,400,225</td>
<td>1,500,400</td>
<td>2,190,225</td>
<td>3,160,100</td>
<td>4,170,225</td>
<td>6,250,225</td>
</tr>
<tr>
<td></td>
<td>4,140,100</td>
<td>1,400,400</td>
<td>8,450,400</td>
<td>7,170,225</td>
<td>6,190,225</td>
<td>5,160,100</td>
<td>2,170,225</td>
<td>3,100,100</td>
</tr>
<tr>
<td></td>
<td>3,120,100</td>
<td>2,400,400</td>
<td>5,190,225</td>
<td>1,080,100</td>
<td>8,450,400</td>
<td>7,090,100</td>
<td>6,400,400</td>
<td>4,160,225</td>
</tr>
<tr>
<td></td>
<td>6,250,225</td>
<td>3,110,100</td>
<td>8,400,400</td>
<td>5,260,225</td>
<td>7,180,100</td>
<td>2,400,400</td>
<td>4,180,225</td>
<td>1,500,400</td>
</tr>
<tr>
<td></td>
<td>8,400,400</td>
<td>1,110,100</td>
<td>7,400,400</td>
<td>2,110,100</td>
<td>5,280,225</td>
<td>6,110,100</td>
<td>4,280,225</td>
<td>3,380,400</td>
</tr>
<tr>
<td></td>
<td>6,140,100</td>
<td>5,280,225</td>
<td>1,380,400</td>
<td>2,100,100</td>
<td>8,400,400</td>
<td>7,160,225</td>
<td>3,180,100</td>
<td>4,500,400</td>
</tr>
</tbody>
</table>

Table 2: Due dates and cost parameters

<table>
<thead>
<tr>
<th>Jobs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c_{ij}^1)</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(c_{ij}^2)</td>
<td>200</td>
<td>120</td>
<td>200</td>
<td>160</td>
<td>200</td>
<td>120</td>
<td>200</td>
<td>160</td>
<td>120</td>
<td>160</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>160</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>(c_{ij}^3)</td>
<td>20000</td>
<td>12000</td>
<td>20000</td>
<td>16000</td>
<td>20000</td>
<td>12000</td>
<td>20000</td>
<td>16000</td>
<td>12000</td>
<td>16000</td>
<td>12000</td>
<td>12000</td>
<td>12000</td>
<td>16000</td>
<td>20000</td>
<td>20000</td>
</tr>
<tr>
<td>(d_{ij})</td>
<td>5000</td>
<td>11000</td>
<td>8000</td>
<td>9000</td>
<td>6000</td>
<td>13000</td>
<td>6000</td>
<td>7000</td>
<td>10000</td>
<td>10000</td>
<td>11000</td>
<td>10000</td>
<td>11000</td>
<td>8000</td>
<td>6000</td>
<td>8000</td>
</tr>
</tbody>
</table>

Table 3: Total cost of our method and decision-making rule with cost parameters [2]. \(S_{\text{ij}}\)'s are set to zero

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Our method</th>
<th>Decision-making rule with cost parameters [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1183721</td>
<td>997250</td>
</tr>
<tr>
<td>Uniform</td>
<td>1481293</td>
<td>1031100</td>
</tr>
<tr>
<td>Exponential</td>
<td>3034265</td>
<td>2375400</td>
</tr>
</tbody>
</table>

In order to obtain meaningful values for \(D_{\text{ij}}\)'s, we first run the simulation model considering \(S_{\text{ij}} = 0\) for all jobs. Then we can calculate the value of standard deviation for the completion time for each job in the \(I_{\text{max}}\) iterations. This standard deviation divided by 10 and rounded up seems to be an appropriate value for \(D_{\text{ij}}\).

In the case of machine breakdowns, we let \(1 - \frac{1}{\lambda_i} = \sum_{j=1}^{N} \tilde{p}_{ij}\) and \(f_i = \left[ \frac{0.5}{\lambda_i} \right]\) to ensure that breakdowns are meaningfully probable to happen in the course of processing any job.

Tables 3, 4, 5 and 6 contain the results and comparisons of our method and the decision-making rule with cost parameters method presented in [2]. According to [2], all the average expenses reported for the decision-making rule with cost parameters method are also drawn from 200 simulation runs.
Table 4: Results of our method and decision-making rule with cost parameters [2]. Process times are Normal

<table>
<thead>
<tr>
<th>Jobs</th>
<th>(t_j)’s of our method</th>
<th>(t_j)’s of the decision-making rule with cost parameters [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>282</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>694</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>543</td>
<td>750</td>
</tr>
<tr>
<td>4</td>
<td>768</td>
<td>900</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>62.5</td>
</tr>
<tr>
<td>6</td>
<td>1431</td>
<td>1800</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>8</td>
<td>121</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>579</td>
<td>1050</td>
</tr>
<tr>
<td>11</td>
<td>1023</td>
<td>900</td>
</tr>
<tr>
<td>12</td>
<td>281</td>
<td>300</td>
</tr>
<tr>
<td>13</td>
<td>1480</td>
<td>1275</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>125</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>62.5</td>
</tr>
<tr>
<td>Total cost</td>
<td>88765</td>
<td>80125</td>
</tr>
</tbody>
</table>

Table 5: Results of our method and decision-making rule with cost parameters [2]. Process times are Uniform

<table>
<thead>
<tr>
<th>Jobs</th>
<th>(t_j)’s of our method</th>
<th>(t_j)’s of the decision-making rule with cost parameters [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>321</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>676</td>
<td>375</td>
</tr>
<tr>
<td>3</td>
<td>459</td>
<td>250</td>
</tr>
<tr>
<td>4</td>
<td>581</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>125</td>
</tr>
<tr>
<td>6</td>
<td>975</td>
<td>675</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>8</td>
<td>38</td>
<td>250</td>
</tr>
<tr>
<td>9</td>
<td>211</td>
<td>450</td>
</tr>
<tr>
<td>10</td>
<td>529</td>
<td>600</td>
</tr>
<tr>
<td>11</td>
<td>410</td>
<td>675</td>
</tr>
<tr>
<td>12</td>
<td>503</td>
<td>600</td>
</tr>
<tr>
<td>13</td>
<td>790</td>
<td>450</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>500</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total cost</td>
<td>133201</td>
<td>97835</td>
</tr>
</tbody>
</table>

6. INTERPRETATION, CONCLUSION AND FUTURE WORKS

This paper considers a stochastic job shop scheduling problem with machine breakdowns. In stochastic job shop, every time a machine is freed, a rule is needed to choose one job from all the jobs which are in line to be passed on the machine. A heuristic rule was devised based on the central limit theorem to deal with this situation. The extensive simulation was used to test our rule. Based on the results of simulation (Tables 3 to 6) the following conclusions can be drawn:

1. Due to higher variance in total cost (which is also reported in [2]), many of \(t_j\)’s are computed to be zero in the case of exponential process times.
Table 6: Results of our method and decision-making rule with cost parameters [2]. Process times are Exponential

| Jobs | S_j^{'s of our method} | S_t^{'s of the decision-making rule with cost parameters [2] | |
|------|------------------------|----------------------------------------------------------|
| 1    | 0                      | 0                                                        |
| 2    | 121                    | 675                                                      |
| 3    | 0                      | 250                                                      |
| 4    | 0                      | 150                                                      |
| 5    | 0                      | 125                                                      |
| 6    | 382                    | 300                                                      |
| 7    | 38                     | 100                                                      |
| 8    | 0                      | 125                                                      |
| 9    | 412                    | 600                                                      |
| 10   | 0                      | 450                                                      |
| 11   | 0                      | 525                                                      |
| 12   | 0                      | 450                                                      |
| 13   | 461                    | 300                                                      |
| 14   | 0                      | 125                                                      |
| 15   | 0                      | 0                                                        |
| 16   | 118                    | 375                                                      |
| Total | 294575                | 213720                                                   |

2. Total cost of shop floor in our stochastic job-shop scheduling problems with machine breakdowns is more than what is reported in [2]. This is clearly because of our assumption about machine breakdowns. In other words, since machine may break during processing jobs, it takes longer for jobs to be finished. According to Table 2 costs of tardiness are much more than inventory costs (this is a realistic assumption). Therefore it can be concluded that higher costs in our problems are due to more tardy costs.

3. Start time of jobs in our method is, for most of jobs, less than what is reported in [2]. This happens because jobs should start sooner in order to compensate for higher tardy costs which are likely to happen due to machine breakdowns.

For future works on stochastic job shop scheduling, one can concentrate on more precise methods of decision making. One may argue that, at any time a competition is raised, the status of all the jobs in the shop floor can affect our decision. However no information was taken on the non-competing jobs in to consideration. This idea is being implemented on stochastic job shop scheduling by using the concept of markov decision process [10].

REFERENCES

An innovative fuzzy time series model for forecasting

S.M.T. Fatemi Ghomi, M.R. Marjani*
Department of Industrial Engineering
Amirkabir University of Technology
Tehran, Iran

Abstract
Fuzzy time series models are major useful tools in forecasting based on vague or incomplete data. In recent years, there have been proposed different models to forecast more accurately. Fuzzification and establishment of fuzzy relations are two major steps of fuzzy time series, which have been touched in previous studies. It has been shown that fuzzification affects forecasting results in fuzzy time series. This study proposes a new method of fuzzy time series based on fuzzification improvement. The method has been applied on forecasting the enrollment of University of Alabama that has been considered in literature. Comparing the forecasted values with the results of prior methods demonstrate that the accuracy of proposed method in forecasting is superior to previous methods.

1. Introduction
Time series models are major useful tools in forecasting. However, if the historical data are very little or linguistic, the statistical methods could not be successful. Since linguistic values can easily be defined into fuzzy sets, fuzzy sets theory was naturally considered to deal with this situation and therefore the concept of fuzzy time series was proposed [1]. Due to its ability in problems with imprecise data, fuzzy time series models have attracted attention of many researchers in recent years. The major objective in this area of forecasting is to develop a method to providing more accuracy in the forecasted values [15].

Different methods have been proposed and studied to solve fuzzy time series problems. Each fuzzy time series essentially consists of three major steps: fuzzification, establishment of fuzzy relationships, and defuzzification. Some studies have focused on the fuzzification step [2 and 3]. Many studies also have focused on the establishment of fuzzy relationships [4-10].

This study focused on the fuzzification part of the fuzzy time series and its aim is to propose a new approach in interval definition. Here the assumption of equal and separate intervals is released and the intervals are defined by clustering the data. The structure of this paper is organized as follows: The last works in this area are reviewed in section 2. In section 3 the concepts of time series and method of interval definition are reviewed. The new method is proposed in section 4. In section 5, the performance of the model in forecasting Alabama University’s enrollments will be compared with other methods. Finally, summary and conclusions will be drawn in Section 6.

2. Literature review
Fuzzy time series was first introduced by Song and Chissom in 1993 [1,4 and 5]. They implemented the developed fuzzy time series models on the historical enrollment data of University of Alabama. Chen by using arithmetic operations presented a simplified method for time series forecasting [6]. Huarng proposed a heuristic model for time series forecasting using heuristic increasing and decreasing relations to improve the forecast of enrollments and also implemented it for TAIEX forecasting [8]. He also presented two approaches (distribution- and average-based) in determining lengths of intervals to improve forecasting with fuzzy time series [2]. Chen presented a new method based on high-order fuzzy time series [7]. He raised some points of ambiguity to the trend in forecast and suggested to use high-order fuzzy logical relationship groups to deal with ambiguity. On comparative study, the third-order

* Corresponding author: E-mail: fatemi@aut.ac.ir (S.M.T. Fatemi Ghomi), marjani@aut.ac.ir (M.R. Marjani)
model was found to have much more accuracy. Song extended the Chen’s model by proposing a method to select the best order for fuzzy time series model; he found that the third order was better than others [11]. Lee and Chou modified the Chen’s method of by redefining the Universe of discourse and subsequent partition of intervals [12]. Also Chen and Hsu proposed a new method in definition of the intervals by dividing the universe of discourse into intervals and then dividing these intervals into four, three, two and one according to the frequency of their occurrences and framed some rules for forecasting [13]. Yu presented a refined fuzzy time series model which can be used to adjust the lengths of the intervals determined during the early stages of forecasting, when the fuzzy relationships are formulated [3]. Yu, proposed weighted models to tackle both recurrence and weighting in fuzzy time series forecasting [10]. She argued that recurrent fuzzy relationships, which were simply ignored in previous studies, and different weights for various fuzzy relationships should be considered in forecasting. Own and Yu presented a heuristic higher-order model by introducing a heuristic function [14]. The trend of time series is used as the heuristic knowledge to specify the difference between times to an increase, a decrease or no change as a parameter. Tsaur et al., used the concept of entropy to measure the degree of fuzziness of a system and to determine a time T of which the data approaches steady state [15]. Huarng and Yu proposed the use of a type 2 fuzzy time series model to utilize more than one variable’s observations in forecasting [9]. In their Type 2 model, extra observations are used to enrich or to refine the fuzzy relationships obtained from Type 1 models and then to improve forecasting performance. Huarng and Yu, applied a back propagation neural network because of its nonlinear structures to forecast fuzzy time series [16]. They proposed two models: a basic model using a neural network approach to forecast all of the observations, and a hybrid model consisting of a neural network approach to forecast the known patterns as well as a simple method to forecast the unknown patterns. Singh, present a new method in fuzzy time series using a difference parameter as fuzzy relation for forecasting. It provides simple computational algorithms of complexity in linear order. It minimizes the time of generating relational equations by using complex min-max composition operations and the time consumed by the various defuzzification process [17]. Li and Cheng, proposes a novel deterministic forecasting model to control the uncertainty in forecasting. In addition, the maximum length of subsequence in a fuzzy time series resulting in a certain state is deterministically quantified [18].

3. REVIEW OF FUZZY TIME SERIES

The concepts of fuzzy time series found in [1–18] are summarized as follows:

Definition 1: Let \( Y(t) \) \((t = \ldots, 0, 1, 2, 3, \ldots)\), is a subset of real numbers, be the Universe of discourse on which fuzzy sets \( f_i(t) \), \((i = 1, 2, \ldots)\) are defined and \( F(t) \) is the collection of \( f_i(t) \), \((i = 1, 2, \ldots)\), then \( F(t) \) is defined as fuzzy time series on \( Y(t) \) \((t = \ldots, 0, 1, 2, 3, \ldots)\).

Definition 2: If \( F(t) \) is caused by \( F(t - 1) \) only, which is denoted by \( F(t - 1) \rightarrow F(t) \), then there exists a fuzzy relationship between \( F(t) \) and \( F(t - 1) \) and can be expressed as \( F(t) = F(t - 1) \times R(t, t - 1) \). Here, “\( X \)” is an operator. The relation \( R \) is called first-order model of \( F(t) \).

Definition 3: Suppose \( F(t) \) \((t = \ldots, 0, 1, 2, 3, \ldots)\), is a fuzzy time series. If for different times \( t_1 \) and \( t_2 \), \( R(t_1, t_1 - 1) = R(t_2, t_2 - 1) \), then \( F(t) \) is called a time invariant fuzzy time series.

Definition 4: If \( F(t) \) is caused by more fuzzy sets, \( F(t - n), F(t - n + 1), \ldots, F(t - 1) \), which is denoted by \( F(t - n), F(t - n + 1), \ldots, F(t - 1) \rightarrow F(t) \), then this relation is called \( n \)th order fuzzy time series model.

Definition 5: If \( F(t) \) is caused by \( F(t - 1), F(t - 2), \ldots, F(t - m) \), \((m > 0)\) simultaneously and the relations are time variant. The \( F(t) \) is said to be time variant fuzzy time series and the relation can be expressed as \( F(t) = F(t - 1) \times R^w(t, t - 1) \) Where, \( w > 1 \) is the number of years before \( t \) various complicated computational methods are available for the computations of the relation \( R^w(t, t - 1) \).

The procedure of fuzzy time series forecasting the enrollments proposed by Song and Chissom includes the following steps [4]:

1. Define the universe of discourse within which the historical data are and upon which the fuzzy sets will be defined,
2. Partition the universe of discourse into several even lengthy intervals,
3. Define the fuzzy sets on the universe,
4. Fuzzify the data, i.e. find out an equivalent fuzzy set to each data,
5. Establish the fuzzy relationships from the historical knowledge,
6- Calculate the Forecast outputs,
7- Defuzzify the forecasting results.

Song and Chissom, time-variant model is similar to their time-invariant model except in step except in Establishing the fuzzy relationships and forecasting, which they choose a model basis w and calculate extrapolation operator RW (t, t - 1) to forecast [5].

Chen proposed a new model, which used arithmetic operation [6]. His model did not need numerous calculations. Chen also proposed a method of forecasting enrollments based on high-order fuzzy time series of various orders: second, third, fourth and fifth, this was appeared in establishment of the relationships [7]. The main structure of all this models is similar. They all have fuzzification the data, establishment the rules for forecasting and defuzzification. Huarng, has showed that in fuzzification process, different lengths of intervals will lead to different forecasting results. However, shorter intervals lead to results that are more accurate [2]. He proposed that the effective lengths of intervals be used. The algorithm for determining the distribution length of intervals is as follows [2]:

1- Calculate all the absolute differences between of data and the average of them.
2- Determine the base for length of intervals according to the average.
3- Plot the cumulative distribution of the first differences.
4- The largest length that is smaller than at least half the first differences is choose as the length of intervals.

Moreover, the Algorithm for average-based length is [2]:

1- Calculate all the absolute differences between of data and the average of them.
2- Take one-half the average (in step 1) as the length.
3- Determine the base for length of intervals according to the length (in step 2).
4- Round the length according to the determined base as the length of intervals.

4. NEW METHOD IN DETERMINING THE LENGTH OF INTERVALS

As it is said before, the fuzzy time series consists of three major processes: fuzzification, the establishment of fuzzy relationships and defuzzification. This study focused on the first one, i.e. it presented a new method in determining of fuzzy sets; although the defuzzification process is also changed. In all last works, there exists a hidden assumption that the intervals must be equal and separate. However, in the fuzzification process each data is defined by more than one interval and by the degree that this data belongs to each interval. While it is true, it can be obviously realized that the aim of these two steps as a whole is to define a certain number of fuzzy sets and determine the membership function of data in each fuzzy set. To do this, the fuzzy clustering can be used simply. Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters) [19]. If fuzzy clustering is used to partitioning the data, there will be fuzzy sets with overlaps.

In this study the fuzzy C-mean algorithm is used for clustering. To do this first the data are sorted in an increasing order and are partitioned in certain number of fuzzy sets. To find out the equivalent fuzzy set of each data, it can be done the same as Song and Chissom, i.e. assume that if the maximum membership of one data is under fuzzy set $A_k$, then that data can be treated as $A_k$ [4]. Afterward, data are returned to prior order with assigned fuzzy sets to each one. It was observed that three or more successive data that was assigned to one fuzzy set in the original sequence may lead to big errors; because they have a same meaning in forecasting. To avoid these errors, it was suggested to increase the number of clusters one by one up to there is not any clusters with three or more successive data.

The algorithm of proposed method is:
1- Define the universe of discourse within which the historical data are.
2- Fuzzification: Sort the historical data in ascending manner and cluster them into certain number of fuzzy sets. Then rearrange the data in their prior sequence. If three or more successive data was assigned to one fuzzy set, then increase the number of clusters by one, and repeat this stage.
3- Establish the fuzzy relations and forecasting.
4- Defuzzification: here the cluster centers can be used simply.
The results of this contribution in both first and high order time series are great. This method is used in forecasting the enrollment of the University of Alabama and the results are compared with the results of other methods. The Fuzzy sets of enrollments data are shown in Table 1. The enrollment data of the University of Alabama and the equivalent fuzzy sets to each year's enrollment are shown in Table 2. The average of interval length in this method is 1081.2.

Table 1: Fuzzy sets of enrollments data from clustering

<table>
<thead>
<tr>
<th>Fuzzy set</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
<th>A11</th>
<th>A12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center</td>
<td>1305</td>
<td>1372</td>
<td>1469</td>
<td>1516</td>
<td>1544</td>
<td>1559</td>
<td>1592</td>
<td>1638</td>
<td>1686</td>
<td>1815</td>
<td>1892</td>
<td>1933</td>
</tr>
</tbody>
</table>

Table 2: Historical and fuzzified enrollments

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual enroll</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
<th>A11</th>
<th>A12</th>
<th>Fuzzified enroll</th>
<th>Forecasted enroll</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>13055</td>
<td>0.99</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A1</td>
<td>-</td>
</tr>
<tr>
<td>1972</td>
<td>13563</td>
<td>0.09</td>
<td>0.87</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A2</td>
<td>13722</td>
</tr>
<tr>
<td>1973</td>
<td>13867</td>
<td>0.03</td>
<td>0.91</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A3</td>
<td>14208.5</td>
</tr>
<tr>
<td>1974</td>
<td>14696</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A4</td>
<td>14208.5</td>
</tr>
<tr>
<td>1975</td>
<td>15460</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.99</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A5</td>
<td>15448</td>
</tr>
<tr>
<td>1976</td>
<td>15311</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.39</td>
<td>0.45</td>
<td>0.11</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A6</td>
<td>15403.33</td>
</tr>
<tr>
<td>1977</td>
<td>15603</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A7</td>
<td>15403.33</td>
</tr>
<tr>
<td>1978</td>
<td>15861</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
<td>0.9</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A8</td>
<td>15925</td>
</tr>
<tr>
<td>1979</td>
<td>16807</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.97</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A9</td>
<td>16862</td>
</tr>
<tr>
<td>1980</td>
<td>16919</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A10</td>
<td>17133.33</td>
</tr>
<tr>
<td>1981</td>
<td>16388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A11</td>
<td>17133.33</td>
</tr>
<tr>
<td>1982</td>
<td>15433</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.99</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A12</td>
<td>15448</td>
</tr>
<tr>
<td>1983</td>
<td>15497</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.78</td>
<td>0.19</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A13</td>
<td>15403.33</td>
</tr>
<tr>
<td>1984</td>
<td>15145</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.99</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A14</td>
<td>15403.33</td>
</tr>
<tr>
<td>1985</td>
<td>15163</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A15</td>
<td>15545</td>
</tr>
<tr>
<td>1986</td>
<td>15984</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.93</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A16</td>
<td>15545</td>
</tr>
<tr>
<td>1987</td>
<td>16859</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A17</td>
<td>16862</td>
</tr>
<tr>
<td>1988</td>
<td>18150</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>A18</td>
<td>17133.33</td>
</tr>
<tr>
<td>1989</td>
<td>18970</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.98</td>
<td>0.02</td>
<td>0</td>
<td>A19</td>
<td>18923</td>
</tr>
<tr>
<td>1990</td>
<td>19328</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>A20</td>
<td>19332</td>
</tr>
<tr>
<td>1991</td>
<td>19337</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>A21</td>
<td>19127.5</td>
</tr>
<tr>
<td>1992</td>
<td>18876</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.98</td>
<td>0.01</td>
<td>A22</td>
<td>19127.5</td>
</tr>
</tbody>
</table>

The fuzzy relations can be obtained from table 2 as follows:

\[ A_1 \rightarrow A_2, A_2 \rightarrow A_3, A_3 \rightarrow A_4, A_4 \rightarrow A_5, A_5 \rightarrow A_6, A_6 \rightarrow A_7, A_7 \rightarrow A_8, A_8 \rightarrow A_9, A_9 \rightarrow A_{10}, A_{10} \rightarrow A_{11}, A_{11} \rightarrow A_{12} \]

From the above relations and centers of fuzzy sets, these values are found out for forecasting:

\[ A_1 \rightarrow 13722, \quad A_2 \rightarrow \frac{13722 + 14695}{2} = 14208.5, \quad A_3 \rightarrow 15448, \quad \ldots \]

The results are shown in Table 2. This fuzzified data can be used in Chen’s high order models [7]. As it was said before, Chen’s comparative study showed that Third-order model lead to best results [7]. The fuzzified data is applied in Chen’s Third-order model. These results are shown in Table 3.
Table 3: Forecasted enrollments based on 1st and 3rd order models

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual enrollment</th>
<th>Fuzzified enrollment</th>
<th>1st order forecast</th>
<th>3rd order relation</th>
<th>3rd order forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>13055</td>
<td>A1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1972</td>
<td>13563</td>
<td>A2</td>
<td>13722</td>
<td>#A1-A2</td>
<td>13722</td>
</tr>
<tr>
<td>1973</td>
<td>13867</td>
<td>A2</td>
<td>14208.5</td>
<td>A1-A2-A2</td>
<td>14695</td>
</tr>
<tr>
<td>1974</td>
<td>14696</td>
<td>A3</td>
<td>14208.5</td>
<td>A2-A2-A3</td>
<td>15448</td>
</tr>
<tr>
<td>1975</td>
<td>15460</td>
<td>A5</td>
<td>15448</td>
<td>A2-A3-A5</td>
<td>15448</td>
</tr>
<tr>
<td>1976</td>
<td>15311</td>
<td>A5</td>
<td>15403.33</td>
<td>A3-A5-A5</td>
<td>15448</td>
</tr>
<tr>
<td>1977</td>
<td>15603</td>
<td>A6</td>
<td>15403.33</td>
<td>A5-A5-A6</td>
<td>15597</td>
</tr>
<tr>
<td>1978</td>
<td>15861</td>
<td>A7</td>
<td>15925</td>
<td>A5-A6-A7</td>
<td>15925</td>
</tr>
<tr>
<td>1979</td>
<td>16807</td>
<td>A9</td>
<td>16862</td>
<td>A6-A7-A9</td>
<td>16862</td>
</tr>
<tr>
<td>1980</td>
<td>16919</td>
<td>A9</td>
<td>17133.33</td>
<td>A7-A9-A9</td>
<td>16862</td>
</tr>
<tr>
<td>1981</td>
<td>16388</td>
<td>A8</td>
<td>17133.33</td>
<td>A9-A9-A8</td>
<td>16388</td>
</tr>
<tr>
<td>1982</td>
<td>15433</td>
<td>A5</td>
<td>15448</td>
<td>A9-A8-A5</td>
<td>15448</td>
</tr>
<tr>
<td>1983</td>
<td>15497</td>
<td>A5</td>
<td>15403.33</td>
<td>A8-A5-A5</td>
<td>15448</td>
</tr>
<tr>
<td>1984</td>
<td>15145</td>
<td>A4</td>
<td>15403.33</td>
<td>A5-A5-A4</td>
<td>15165</td>
</tr>
<tr>
<td>1985</td>
<td>15163</td>
<td>A4</td>
<td>15545</td>
<td>A5-A4-A4</td>
<td>15165</td>
</tr>
<tr>
<td>1986</td>
<td>15984</td>
<td>A7</td>
<td>15545</td>
<td>A4-A4-A7</td>
<td>15925</td>
</tr>
<tr>
<td>1987</td>
<td>16859</td>
<td>A9</td>
<td>16862</td>
<td>A4-A7-A9</td>
<td>16862</td>
</tr>
<tr>
<td>1988</td>
<td>18150</td>
<td>A10</td>
<td>17133.33</td>
<td>A7-A9-A10</td>
<td>18150</td>
</tr>
<tr>
<td>1989</td>
<td>18970</td>
<td>A11</td>
<td>18923</td>
<td>A9-A10-A11</td>
<td>18923</td>
</tr>
<tr>
<td>1990</td>
<td>19328</td>
<td>A12</td>
<td>19332</td>
<td>A10-A11-A12</td>
<td>19332</td>
</tr>
<tr>
<td>1991</td>
<td>19337</td>
<td>A12</td>
<td>19127.5</td>
<td>A11-A12-A12</td>
<td>19332</td>
</tr>
<tr>
<td>1992</td>
<td>18876</td>
<td>A11</td>
<td>19127.5</td>
<td>A12-A12-A11</td>
<td>18923</td>
</tr>
</tbody>
</table>

Figure 1: Actual enrollments vs. Forecasted enrollments

Also the proposed method is being implemented for forecasting the Wheat production. The historical time series data of Wheat production are from reference [17]. The historical time series data of wheat production is in terms of productivity in kg per hectare. The proposed method has been implemented and the results are shown in tables 4 and 5.

Table 4: Fuzzy sets of wheat production from clustering

<table>
<thead>
<tr>
<th>Fuzzy set</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center</td>
<td>3224.2</td>
<td>3406.3</td>
<td>3569.6</td>
<td>3683.7</td>
<td>3749.1</td>
<td>3862.8</td>
<td>3928</td>
<td>4174.1</td>
<td>4289.1</td>
<td>4511.5</td>
</tr>
</tbody>
</table>
5. PERFORMANCE COMPARISON

To measure the accuracy of forecasting the Mean Square Error (MSE) can be calculated as follows:

\[
\text{Mean Square Error (MSE)} = \frac{1}{n} \sum_{i=1}^{n} (Actual_i - Forecast_i)^2
\]

In this section, the forecasting performance of the proposed fuzzy time series models on the University of Alabama enrollments data from 1971 to 1992 is compared with the previous models, which all used these data. The MSEs of their results are summarized in Table 6. The comparison of the MSE of forecasting in different models showed that the proposed models has the best results. It must be noticed that the average length of intervals determined by proposed method is 1081.2, which is a little more than 1000.

Table 6: MSE of forecasting enrollments in different methods

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
<th>Method</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song &amp; Chissom time invariant</td>
<td>458437.5</td>
<td>Singh</td>
<td>133700.4</td>
</tr>
<tr>
<td>Song &amp; Chissom time variant</td>
<td>775686.8</td>
<td>Huarng 1000 length of intervals</td>
<td>407521</td>
</tr>
<tr>
<td>Chen’s 1st order</td>
<td>439420.8</td>
<td>Huarng 200 length of intervals</td>
<td>104640</td>
</tr>
<tr>
<td>Chen’s 3rd order</td>
<td>86694</td>
<td>Li &amp; Cheng 1000 length of intervals</td>
<td>85040</td>
</tr>
<tr>
<td>Huarng heuristic</td>
<td>239483.1</td>
<td>Li &amp; Cheng 600 length of intervals</td>
<td>26354</td>
</tr>
<tr>
<td>Lee and Chou</td>
<td>240047</td>
<td>Proposed 1st order</td>
<td>123525.9</td>
</tr>
<tr>
<td>Tsaur et al.</td>
<td>138322.4</td>
<td>Proposed 3rd order</td>
<td>3066.2</td>
</tr>
</tbody>
</table>

Also the performance of the proposed models are compared with other models in forecasting of wheat production from 1982 to 2000 and the results are summarized in table 7. The comparative study of MSE, as shown in Table 6 and table 7 shows that the forecast by the proposed methods are of higher accuracy over the other methods.
6. DISCUSSION

This study focused on both fuzzification and defuzzification processes of fuzzy time series. In this study an innovative method is presented to determine fuzzy sets. In all last works, there exists an assumption that the intervals of defined fuzzy sets must be equal and separate. Albeit in the fuzzification process each data is defined by degree of membership to each fuzzy set. Here, unlike last studies in this field, the fuzzy clustering has been used simply. This method led to unequal fuzzy sets with overlaps. The centers of these sets, which are used in forecasting process, have much more accuracy than others from previous methods. The comparative results confirm this. The number of fuzzy sets has a significant role in precision of forecasting results. Proposed method of determining the number of clusters is another effective aspect of this study.

7. SUMMERY

This paper proposed a new approach in fuzzy time series based forecasting. The fuzzy c-mean algorithm is used for clustering ordered data in certain number of fuzzy sets to find out best partitioning of data. In forecasting problems with limit or incomplete data, this model can reach to assured results. The proposed method, which has been implemented in both first and third order fuzzy time series models, led to obtaining satisfactory results. The new model has been applied on forecasting the enrollments of University of Alabama and wheat production to have a comparative study with the existing methods and outperform them with the least mean square errors. Thus, it can be applied to improve the forecasting based on fuzzy time series.

REFERENCES


A new approach in performance evaluation with Artificial Neural Networks

Abdolhamid Safaei Ghadikolaei*, Ali Zareshahi

Department Of Industrial Management
The Mazandaran University
Mazandaran, Babolsar, Iran

ABSTRACT

There has been a long tradition in business and economics to use frontier analysis to assess a production unit’s performance. The first attempt utilized the data envelopment analyses (DEA) which use a mathematical programming approach. Recently someone use artificial neural networks (ANNs) for performance evaluation and show that ANNs are a promising alternative to traditional approaches, to approximate production functions more accurately and measure efficiency and productivity under non-linear contexts, with minimum assumption. Although ANNs have many priority on other performance evaluation methods, but when a Decision Making Unit (DMU) have multi outputs ANNs only offer multi efficiency score for each output and thereupon the final ranking for a DMU is not met. Therefore this paper presents a new approach in measuring efficiency by use of artificial neural network. According to this method every DMU and every output efficiency assumed as an alternative and attribute respectively. Finally Analytic Hierarchy Process (AHP) is applied to transform multi ranking to a single one for each DMU. For more perception, efficiency of 66 branches of Saderat bank in Iran was evaluated and results compared with traditional DEA models and DEA-ANN. results show that, every three methods have same trends in ranking and proposal method has satisfactory correlations with other perspectives.

1. INTRODUCTION

Productivity and efficiency in converting resource (inputs) into good and service (outputs) have been key issues in private and public sectors. There have been countless articles of efficiency analysis in the literature of operation research and econometrics. Efficiency frontier analysis has been an important approach of evaluating firms’ performance in private and public sectors [1]. Five different approaches, namely, data envelopment analysis (DEA), free disposal hull (FDH), stochastic frontier approach (SFA), also called econometric frontier approach (EFA), thick frontier approach (TFA), and distribution free approach (DFA), have been reported in the literature as frontier analysis methods to evaluate a decision making unit (DMU) efficiency. These approaches primarily differ in how much restriction is imposed on the specification of the best practice frontier and the assumption on random error and inefficiency [2].

Two first approaches mentioned above are non-parametric and remains are parametric approaches. The parametric frontier method assumes a particular functional form for the efficiency boundary. A procedure that is called corrected OLS (COLS) is used to measure technical inefficiency. For this purpose First an average practice frontier is estimated using OLS. Then this frontier is corrected by shifting the intercept until all residuals except one become negative. Hereby primitive average frontier (OLS) transformed to an envelopment frontier (COLS). Versus, non-parametric methods do not assume a particular production function. In this approaches a piece-wise linear frontier is directly constructed from the observations themselves by applying mathematical programming techniques [3]. Every one of these two approaches has its strength and weakness. Hence, recently some of researchers attempt to find a method to corporate two approaches strength. In this way, Reference [4] for the first introduces Artificial Neural Networks (ANN) as a non-parametric method for performance evaluation. Their application is bank with multi output: 4 inputs, 3 outputs. In this study researchers compare ANNs and DEA. In a simulation, they conclude that DEA is superior to the ANNs for measurement purposes and ANNs are similar to DEA for ranking units. Reference [5] analyses the London underground efficiency with time series data for 1970-1994 where there are 2 inputs: fleet and workers and 1 output: kms. They explain how the ANNs result similar to Corrected Ordinary Least Squares (COLS) and DEA, but ANNs offer advantages at the decision making, the impact of constant vs. variable returns to scale or congestion areas.

* Corresponding author: Tel.: (098) 112-5243001; Fax: (098) 112-5242705; E-mail: ab.safaei@umz.ac.ir
Reference [6] use a neural network for a simulated non-linear production function and compare its performance with traditional alternatives like stochastic frontier and DEA in different observations number and noise scenarios. Their results show that ANNs obtains robust estimations with few variations respecting true efficiency over number of DMUs and injected noise. MLP is superior to traditional techniques when underlying technology is under moderate noise together with more DMUs.

In another study Reference [7] compares some of parametric and non-parametric models by use of ANNs. In this study sample was 72 Spanish municipalities. The considered output is the solid waste and the inputs included the containers, vehicles and worked hours. Stochastic and deterministic Cobb-Douglas were two parametric models and FDH, DEACrs and DEAVrs were three non-parametric models that used in this study. Finding show that although several difference in the quantitative measures are evident, but there are a common trends in results so that the most and the lowest efficient units were identified by all the approaches.

Reference [8] use ANNs for measuring efficiency and productivity growth for seven East Asian economies at manufacturing level, for the period 1963-1998, and relevant comparisons are carried out between DEA and ANN, and SFA and ANN in order to test the ability of ANNs to assess the performance of production units. the sample in this study include 28 manufacturing sectors with 1 output-value added and 2 input-fixed capital formation and employment measured in numbers. Growth index decomposition shows that East Asian economic growth in the future will mainly rely on innovation that is technological progress.

All studies mentioned above, have a similar weakness that is their inability in encountering multi outputs. As observable above, References [5,7,8] use a model with 1 output because ANN calculate one efficiency for each output, therefore in the case that there are multiple outputs, management has multi score as efficiency for each DMU.Obviously in this instance management can not have a comprehensive ranking of DMUs. Then this paper uses AHP to propose a solution to corporate the output efficiency score calculated for each DMU. Proposal method does not require value judgment and therefore is an innovative solution.

Reference [2] proposes another school in performance evaluation by ANNs. They presented a DEA-ANN study for performance assessment of branches of a large Canadian bank. The results are operable to the normal DEA results on the whole. They concluded that the DEA-ANN approach produces a more robust frontier and identifies more efficient units because better performance patterns are explored. Furthermore, for worse performers, it provides the guidance on how to improve their performance to different efficiency ratings. Ultimately, they concluded the neural network approach requires no assumptions about the production function (the major drawback of the parametric approach) and it is highly flexible.

2. ANALYTICAL METHODOLOGY

This study use DEA, ANNs and DEA-ANN as three frontier analysis methods in performance evaluation.

2.1. DATA ENVOLPMENT ANALYSIS

Data Envelopment Analysis (DEA) is a mathematical model that measure the relative efficiency of decision making units with multiple inputs and outputs but with no obvious production function to aggregate the data in its entirety. Relative efficiency is defined as the ratio of total weighted output to total weighted input [9].

Let us assume a set of $j = \{1, \ldots, n\}$ DMUs to be evaluated. Each DMU consumes varied amounts of $m$ different inputs to produce $s$ different outputs. Specially, DMU $j$ consumes amount $x_{ij}$ of inputs ($i=1\ldots m$) and produce amounts $y_{rj}$ of outputs ($r=1\ldots s$). The efficiency of unit $k$ under the assumption of constant and variable returns to scale is given by the solution to the optimization problem in (1) and (2), respectively (Table 1). The formulation of these models has been proposed by [10] and [11]: Where $v_i$ is weight factor for input $i$ estimated by two models, $u_r$ is weight factor for output $r$ estimated by two models, $\omega$ is a sign free variable that characterizes the variable returns to scale model, $\epsilon$ is a non-Archimedean small positive number. Efficiency of unit $k$ is obtained under CRS as $E_k^{CRS} = \frac{1}{\sum_{i=1}^{m} v_i^* x_{ik}}$ and under VRS as $E_k^{VRS} = \frac{1}{\sum_{i=1}^{m} v_i^* x_{ik} - \omega}$ (where the asterisks indicate optimal values).
Table 1: DEA models for performance evaluation

<table>
<thead>
<tr>
<th>Constant Return to Scale (CRS)</th>
<th>Variable Return to Scale (VRS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min $\sum_{i=1}^{m} v_i x_{ik}$</td>
<td>Min $\sum_{i=1}^{m} v_i x_{ik} - \omega$</td>
</tr>
<tr>
<td>S.t. $\sum_{r=1}^{s} u_r y_{rk} = 1$</td>
<td>S.t. $\sum_{r=1}^{s} u_r y_{rk} = 1$</td>
</tr>
<tr>
<td>$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0, \forall j$</td>
<td>$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} - \omega \geq 0, \forall j$</td>
</tr>
<tr>
<td>$v_i, u_r \geq \epsilon$</td>
<td>$v_i, u_r \geq \epsilon$ and $\omega$ unrestricted</td>
</tr>
</tbody>
</table>

2.2. ARTIFICIAL NEURAL NETWORKS

In this section we give a brief review of neural network mechanisms and its applications in efficiency measurement.

2.2.1. A BRIEF REVIEW OF ARTIFICIAL NEURAL NETWORKS

An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. ANNs have been applied when there is no theoretical evidence about the functional forms [12]. The Multi Layer Perceptron (MLP) is perhaps the most popular and widely applied of the many existing ANN types. Reference [13] has shown that subject to mild regularity conditions these models can approximate any function and its derivatives to any degree of accuracy. The MLP basic properties can be summarized in the triplet: multi-layer, feed-forward and supervised neural network. Its processing elements, known as ‘neurons’, are organized in at least three layers: the input layer, the output layer and the hidden layer(s) in between. These neurons are all fully connected between adjacent layers. MLPs are feed-forward networks, i.e. all connections point in one direction, from the input towards the output layer. Finally, they are supervised networks since all patterns of both inputs and outputs must be provided. The development of a neural network model requires the specification of a ‘network topology’ and a ‘training strategy’. The MLP topology is defined by the number of input and output neurons, the number of hidden layers and the number of neurons in each hidden layer. The training strategy refers to the procedure and technique used in estimating the optimal network parameters (known as ‘weights’) [5]. Training involves repeatedly presenting the data to the network. Learning is achieved in the network by altering the values of weighted connections between neurons to bring the output of the network closer to the desired target value. The overall aim is to reduce the Mean Squared-Error (MSE) for the training data. The errors between output and target values are propagated back through the network to attribute them to the weights in the network, these are then altered using the steepest descent method, which aims to reduce the MSE by following the steepest gradient on the error surface [4]. In a MLP, the output is function of the linear combination of hidden units activation, each of one is a non-linear function of the weighted sum of inputs. In this way, from equation 3:

$$y_j = f(x_j; \theta) + \epsilon$$

(3)

Where $x_j$ is the vector of explanatory variables, $\epsilon$ is the error component, $f(x_j; \theta) = \hat{y}$ is the unknown function to estimate from the available information, the network consist of equation 4:

$$\hat{y} = f(\nu_0 + \sum_{j=1}^{m} H(\lambda_j + \sum_{i=1}^{n} x_i w_{ij})v_j)$$

(4)

Where $\hat{y}$: Network output, $F$: Output layer activation function, $H$: Hidden layer activation function, $n$: Number of input units, $m$: Number of hidden units, $x$: Input vector ($i=1,\ldots,n$), $\theta$: Weight vector parameters, $\nu_0$: Output bias,
$\lambda_j$: Hidden unit bias ($j=1\ldots m$), $w_{ij}$: Weight from input unit $i$ to hidden unit $j$, $v_j$: Weight from hidden unit $j$ to output [6].

The activation function for output layer $F(x, w)$ might be a linear function of the inputs through the weights ($w$) plus a constant. After neural training (training set), new observations (validation and/or test sets) are presented to the network to verify the so-called generalization capability. Here it is relevant the statistical classical bias–variance dilemma [14] or overfitting problem [8].

2.2.2. ARTIFICIAL NEURAL NETWORKS AND EFFICIENCY MEASUREMENT

There are several approaches for performance measurement with ANNs. The neural network efficiency will be determined using the predicted values obtained from the solution to the model. Three basics approach herein are following:

1- An approach similar to ordinary least squares (OLS): In this approach the efficiency measure of a unit is established in relation to the average performance. Thus, the indicators will be superior to 1 or 100% when the unit behaves better than average and inferior to 1 or 100% if the unit is “inefficient”. These measures are not directly comparable with the traditional techniques. Reference [4] called this option “non-standardized efficiency” (Equation 6).

$$E^E_i = \frac{y_i}{\bar{y}_i} \quad E^E_i \text{ or } E^\text{MLPMAX}_i = \frac{y_i}{\bar{y}_i + \text{avg}(e_i)} \quad E^E_i = \frac{y_i}{\bar{y}_i + \text{avg}(e_i)}$$

To achieve a real production frontier there are some alternatives:

2- Shift the network by largest positive error: This option is similar to corrected ordinary least squares (COLS). The correction by the largest positive error is sensitive to outliers and the frontier will be deterministic. The efficiency scores take values between 0 and 1. This maximum score is assigned to the unit used for the correction. Reference [4] called this second measure “standardized efficiency” $E^E$ and [15] called this $E^\text{MLPMAX}$ (Equation 7).

In relation to DEA, the non-standardized efficiency ($E^\text{NE}$) tends to overestimate the indicators, and the standardized efficiency ($E^E$) tends to underestimate the measures.

3- Shift the network by a mean of the largest positive errors: this option is similar to modified ordinary least squares (MOLS). In this approach it is possible for a DMU to take efficiency greater than 1 (Equation 8).
2.3. DATA ENVELOPMENT ANALYSIS-ARTIFICIAL NEURAL NETWORKS

Reference [2] integrate Data Envelopment Analysis and Artificial Neural Networks and propose DEA-ANN as an algorithm that explained in follow.

2.3.1. ALGORITHM

1. Collect a data set S which describes the input-output relationship for DMUs. Obtain the preprocessed data set SS after the data in S are divided by an integer 1000 for the purpose of neural network training and simulation.

2. CCR method is used to calculate efficiency score of DMUs in S. The data set SS is grouped into four categories S1, S2, S3, and S4 based on the efficiency scores. The efficiency score intervals of S1 $\epsilon(0.98, 1]$ are referred as ‘strong relative efficient interval’. The efficiency score intervals of S2 $\epsilon(0.8, 0.98]$ are referred as ‘relative efficient interval’. The efficiency score intervals of S3 $\epsilon(0.5, 0.8]$ are referred as ‘relative inefficient interval’ and the efficiencies of S4 $\epsilon(0, 0.5]$ are referred as ‘very inefficient interval’.

3. Train neural network NN1 with S1 and any other two groups of data subset (e.g. S1$^\text{c}$S2$^\text{c}$S3 or S1$^\text{c}$S2$^\text{c}$S4 or S1$^\text{c}$S3$^\text{c}$S4). If the pre-specified epochs or accuracy is satisfied, STOP; go to Step 5. Otherwise, change one training subset and go to Step 3.

4. Apply the trained neural network model to the data set SS to calculate efficiency scores of all DMUs.

5. Post process the calculated efficiency scores by regress analysis between DEA-NN results and CCR DEA results. Do statistics analysis of DEA-NN efficiencies.

3. PERFORMANCE ASSESSMENT OF SADERAT BANK BRANCHES

1. Determination input ($x_i$) and output ($y_i$) variables of the model where $i=1,\ldots,n$ and $k=1,\ldots,l$

2. Collecting data set S in all available previous periods which describes the input–output periods which describes the input–output relationship for DMUs. Assume that there are a set of $j=\{1,\ldots,m\}$ DMUs to be evaluated. Note that the current period data ($S_c$) which evaluation will have been done on them, do not belong to S.

3. Divide S in to two subsets: training ($S_1$) and test ($S_2$) data.

4. Use ANN method to estimate relation between input ($x_i$) and output ($y_i$). For this reason follow these steps:
   - Select architecture and training parameters.
   - Train the model using the training data ($S_1$).
   - Evaluate the model using the test data ($S_2$).
   - Repeat these steps using different architectures and training parameters.
   - Select the best network architecture (ANN*) from the testing data error.

5. Run ANN* for $S_c$.

6. Calculate the error between the real output ($y_i$) and ANN model output ($\hat{y}_i$) in the period which you want to assess the efficiency of its DMUs ($S_c$):

   $\varepsilon_i = y_i - \hat{y}_i \quad i=1,\ldots,n \quad (9)$

7. For $DMU_j$, Measure the efficiency of outputs ($y_i$) with equitation (7).

8. Integrate the efficiency of outputs with AHP and acquire a single ranking for each DMU. Accordingly in follow a proposal algorithm was discussed

3.1. PROPOSAL ALGORITHM

- Assume outputs efficiency and DMUs as attributes and alternatives respectively.
- Plan a hierarchical structure similar figure 2. There are $m$ attributes and $n$ alternatives in the first level respectively.
Calculate a geometric mean \( (a_i) \) for output efficiency (equation 10).

\[
a_i = \left( \prod_{j=1}^{n} E_{y_{ij}} \right)^{\frac{1}{n}} \text{ where } i = 1, \ldots, m
\]  

Prepare Pairwise comparison matrices for attributes level (figure 3). It is obvious that proposal algorithm don’t require to judgments of decision makers.

\[
\begin{array}{ccccccc}
E_{y1} & E_{y2} & E_{y3} & \ldots & E_{ym} \\
E_{y1} & 1 & a_{1}/a_{2} & a_{1}/a_{3} & \ldots & a_{1}/a_{m} \\
E_{y2} & a_{2}/a_{1} & 1 & a_{2}/a_{3} & \ldots & a_{2}/a_{m} \\
E_{y3} & a_{3}/a_{1} & a_{3}/a_{2} & 1 & \ldots & a_{3}/a_{m} \\
& \vdots & \vdots & \vdots & \ddots & \vdots \\
E_{ym} & a_{m}/a_{1} & a_{m}/a_{2} & a_{m}/a_{3} & \ldots & 1 \\
\end{array}
\]

Prepare Pairwise comparison matrices for alternatives level.

Integrate two matrices and acquire final ranking of DMUs.

### 3.2. Case Study

In this study inputs and outputs are employees, terminals, capital and deposits, loans, profit respectively. Calculations related to DEA models and ANNs performed by use of EMS and MATLAB softwares. Final results are shown in table 2. This table shows efficiency scores and ranking earned by each method. Also it is obvious in figure 4 that every three methods have same trends in ranking.
A new approach in performance evaluation with Artificial Neural Networks

Table 2: ranking and efficiency of three models

<table>
<thead>
<tr>
<th>DMU</th>
<th>DEA-CCR</th>
<th>DEA-BCC</th>
<th>DEA-ANN</th>
<th>ANN-AHP</th>
<th>DMU</th>
<th>DEA-CCR</th>
<th>DEA-BCC</th>
<th>DEA-ANN</th>
<th>ANN-AHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>S</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>0.70</td>
<td>36</td>
<td>0.76</td>
<td>16</td>
<td>0.71</td>
<td>11</td>
<td>34</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>0.67</td>
<td>1</td>
<td>1.00</td>
<td>19</td>
<td>0.67</td>
<td>12</td>
<td>35</td>
<td>0.68</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>0.67</td>
<td>36</td>
<td>0.76</td>
<td>29</td>
<td>0.50</td>
<td>20</td>
<td>36</td>
<td>0.70</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>0.86</td>
<td>1</td>
<td>1.00</td>
<td>23</td>
<td>0.58</td>
<td>3</td>
<td>37</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>0.75</td>
<td>11</td>
<td>0.80</td>
<td>10</td>
<td>0.79</td>
<td>11</td>
<td>38</td>
<td>0.66</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0.91</td>
<td>1</td>
<td>1.00</td>
<td>17</td>
<td>0.69</td>
<td>5</td>
<td>39</td>
<td>0.89</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0.85</td>
<td>7</td>
<td>0.88</td>
<td>16</td>
<td>0.71</td>
<td>14</td>
<td>40</td>
<td>0.71</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1.00</td>
<td>1</td>
<td>1.00</td>
<td>16</td>
<td>0.71</td>
<td>8</td>
<td>41</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>0.53</td>
<td>50</td>
<td>0.53</td>
<td>24</td>
<td>0.57</td>
<td>21</td>
<td>42</td>
<td>0.77</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>0.52</td>
<td>45</td>
<td>0.61</td>
<td>30</td>
<td>0.47</td>
<td>12</td>
<td>43</td>
<td>0.59</td>
</tr>
<tr>
<td>11</td>
<td>31</td>
<td>0.50</td>
<td>48</td>
<td>0.58</td>
<td>30</td>
<td>0.47</td>
<td>25</td>
<td>44</td>
<td>0.98</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>0.70</td>
<td>3</td>
<td>0.95</td>
<td>22</td>
<td>0.59</td>
<td>9</td>
<td>45</td>
<td>1.00</td>
</tr>
<tr>
<td>13</td>
<td>20</td>
<td>0.65</td>
<td>42</td>
<td>0.66</td>
<td>21</td>
<td>0.61</td>
<td>22</td>
<td>46</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>31</td>
<td>0.50</td>
<td>51</td>
<td>0.51</td>
<td>28</td>
<td>0.51</td>
<td>14</td>
<td>47</td>
<td>0.95</td>
</tr>
<tr>
<td>15</td>
<td>24</td>
<td>0.60</td>
<td>45</td>
<td>0.61</td>
<td>28</td>
<td>0.51</td>
<td>19</td>
<td>48</td>
<td>1.00</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>0.63</td>
<td>42</td>
<td>0.66</td>
<td>26</td>
<td>0.54</td>
<td>11</td>
<td>49</td>
<td>0.87</td>
</tr>
<tr>
<td>17</td>
<td>15</td>
<td>0.71</td>
<td>6</td>
<td>0.80</td>
<td>21</td>
<td>0.61</td>
<td>10</td>
<td>50</td>
<td>0.52</td>
</tr>
<tr>
<td>18</td>
<td>21</td>
<td>0.63</td>
<td>1</td>
<td>1.00</td>
<td>29</td>
<td>0.50</td>
<td>6</td>
<td>51</td>
<td>0.30</td>
</tr>
<tr>
<td>19</td>
<td>8</td>
<td>0.87</td>
<td>1</td>
<td>1.00</td>
<td>14</td>
<td>0.73</td>
<td>7</td>
<td>52</td>
<td>0.88</td>
</tr>
<tr>
<td>20</td>
<td>26</td>
<td>0.57</td>
<td>1</td>
<td>1.00</td>
<td>24</td>
<td>0.57</td>
<td>11</td>
<td>53</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>12</td>
<td>0.79</td>
<td>5</td>
<td>0.92</td>
<td>25</td>
<td>0.55</td>
<td>4</td>
<td>54</td>
<td>0.50</td>
</tr>
<tr>
<td>22</td>
<td>28</td>
<td>0.55</td>
<td>46</td>
<td>0.60</td>
<td>32</td>
<td>0.45</td>
<td>23</td>
<td>55</td>
<td>0.55</td>
</tr>
<tr>
<td>23</td>
<td>27</td>
<td>0.56</td>
<td>1</td>
<td>1.00</td>
<td>29</td>
<td>0.50</td>
<td>11</td>
<td>56</td>
<td>0.96</td>
</tr>
<tr>
<td>24</td>
<td>32</td>
<td>0.49</td>
<td>49</td>
<td>0.57</td>
<td>31</td>
<td>0.46</td>
<td>24</td>
<td>57</td>
<td>0.59</td>
</tr>
<tr>
<td>25</td>
<td>28</td>
<td>0.55</td>
<td>35</td>
<td>0.78</td>
<td>25</td>
<td>0.55</td>
<td>15</td>
<td>58</td>
<td>0.85</td>
</tr>
<tr>
<td>26</td>
<td>23</td>
<td>0.61</td>
<td>44</td>
<td>0.62</td>
<td>14</td>
<td>0.73</td>
<td>14</td>
<td>59</td>
<td>0.59</td>
</tr>
<tr>
<td>27</td>
<td>25</td>
<td>0.59</td>
<td>46</td>
<td>0.60</td>
<td>34</td>
<td>0.31</td>
<td>24</td>
<td>60</td>
<td>0.56</td>
</tr>
<tr>
<td>28</td>
<td>22</td>
<td>0.62</td>
<td>11</td>
<td>0.80</td>
<td>14</td>
<td>0.73</td>
<td>15</td>
<td>61</td>
<td>0.88</td>
</tr>
<tr>
<td>29</td>
<td>19</td>
<td>0.66</td>
<td>39</td>
<td>0.70</td>
<td>11</td>
<td>0.76</td>
<td>13</td>
<td>62</td>
<td>0.56</td>
</tr>
<tr>
<td>30</td>
<td>16</td>
<td>0.70</td>
<td>8</td>
<td>0.86</td>
<td>11</td>
<td>0.76</td>
<td>13</td>
<td>63</td>
<td>0.52</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1.00</td>
<td>1</td>
<td>1.00</td>
<td>1</td>
<td>1.00</td>
<td>2</td>
<td>64</td>
<td>0.95</td>
</tr>
<tr>
<td>32</td>
<td>17</td>
<td>0.68</td>
<td>38</td>
<td>0.73</td>
<td>12</td>
<td>0.75</td>
<td>14</td>
<td>65</td>
<td>0.43</td>
</tr>
<tr>
<td>33</td>
<td>16</td>
<td>0.70</td>
<td>10</td>
<td>0.82</td>
<td>12</td>
<td>0.75</td>
<td>14</td>
<td>66</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Note: in above table, R and S are Ranking and efficiency score respectively.

Figure 4: Compare ranking of DMUs by each method
4. SUMMARY AND CONCLUSION

This paper analyses the efficiency of 66 Saderat bank branches in Iran by use of DEA, ANNs-DEA and a proposal algorithm namely ANN-AHP. Table 3 show spearman correlation coefficients between DMU ranks in different methods. Results show that proposal method (ANN-AHP) has satisfactory correlations with other method therefore ANN-AHP can introduced as an alternative approach in performance evaluation.

Table 3: spearman rank correlations for the efficiency of each output

<table>
<thead>
<tr>
<th></th>
<th>BCC</th>
<th>DEA-ANN</th>
<th>ANN-AHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR</td>
<td>0.75</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>BCC</td>
<td>0.51</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>DEA-ANN</td>
<td></td>
<td></td>
<td>0.40</td>
</tr>
</tbody>
</table>

REFERENCES

Design of a Framework for Inventory Control – Evaluation of Forecasting and Inventory Control Systems

Philip Hedenstierna, Per Hilletofth*, and Olli-Pekka Hilmola
Logistics Research Centre
School of Technology and Society
University of Skövde
541 28 Skövde, Sweden

ABSTRACT
Managing inventories so that overall costs are kept low, while service levels are maintained is the central issue of inventory control, which only regulates two things: the size and the timing of orders. This is typically executed through a planning method, such as the reorder point system or, less frequently, the periodic order quantity system. These take into account a forecast, supposed to gauge the average future demand, and a predetermined safety stock, buffering against forecast errors and demand uncertainty. Pure demand also influences the system, as transactions affect the inventory level. It is crucial to understand how a complete system of demand, forecasts, safety stock calculations and planning methods work together to measure service level and overall cost of the system. This paper outlines a framework for the unambiguous representation of the relations between methods that interpret environmental parameters to plan orders. A number of simulations based on the framework are run to show, how the integration of the inventory control functions may affect the overall performance of the system. The usefulness of the framework lies in its ability to make a system duplicable (i.e. to transfer an inventory control system to a simulation model, or vice versa). Not only is this property important for creating simulation models that exactly depict the system being analysed, it also enables the study of a complete system for order planning, as opposed to optimising individual methods. Studying an inclusive system allow the same metrics to be used to evaluate changes to any method in the system. Another benefit of this approach is that the system’s metrics directly reflect changes in the environment. Simulations based on this framework are precise and substantially easier to evaluate than models not adhering to any standard.

1. INTRODUCTION
The purpose of inventory control is to ensure service to processes or customers in a cost-efficient manner, which means that the cost of materials acquisition is balanced with the cost of holding inventory [1]. This balancing is done by interpreting data describing the planning environment, i.e. the parameters that may affect the decision, to generate replenishment order times and quantities [2]. The performance of an inventory control system may then be measured by the service and the total cost it generates when applied in a certain environment. Inventory control methods may be classified by whether they determine ordering timing, quantity, or both [3]. For systems that determine only one aspect, such as the reorder point system or the periodic ordering system, the aspect not determined by the system must be calculated beforehand, typically using the economic order quantity or the corresponding inventory cycle time, which enables the generation of order timing and quantity [3]. The parameters that inventory control systems are based on are a forecast of demand, expected lead times, holding costs and ordering costs. Of these, the forecast is of special concern, as it is not a part of the planning environment, but a product thereof. This makes forecasting, given our definition of inventory control, an integral part of the inventory control system. To maintain service when there is forecast and lead time variability, a safety stock mechanism is used, which is based on measures of variability and on the uncertain times that come with the used inventory control model [4]. As the safety stock incurs a holding cost, it may be argued that it should be part of the timing/sizing decision, but it is usually excluded as the cost savings gained from optimisation generally are insignificant [5].

* Corresponding author: Tel.: +46 (0)500 44 85 88; Fax: +46 (0)500 44 87 99; E-mail: per.hilletofth@his.se
We have now discussed three interdependent areas that are usually treated as individual entities. They are all part of generating replenishment orders by interpreting the planning environment, and may consequently affect the performance measures of the system. The current approach to inventory control does generally not consider it as a single system, but as separate methods (see e.g. Axsäter [5], Waters [4], Mattsson & Jonsson [3] and Vollman et.al [6]. An exception is Higgins [7], who describes inventory control as a process (see Figure 1), but does not detail show information flows through the model, nor does he isolate the functions of forecasting, safety stock sizing and inventory control.

![Diagram of the inventory control process](image)

**Figure 1. Higgins’ model of the inventory control process** [7].

Looking at Higgins model, it is easy to realise that corruption of data occurring in an operation or between operations will cause the incorrect data to affect the subsequent operations (see Ashby [8]). When theoretical models are applied to scenarios that follow the assumptions of the models, this is not an issue; but when a model is applied to a scenario it is not designed for, data corruption ensues. Applied to inventory control, this may mean that a simple forecast or a simple inventory control method is applied in an environment that does not reflect the method’s assumptions. When a method’s assumptions are unmet, its performance may be difficult to predict. The scenario of using theoretically improper methods is not unlikely, as businesses may want to utilise inventory control methods that are simple to manage, such as the reorder point system, even when the environment would require a dynamic lot-sizing method such as the Wagner-Whitin, part-period or the Silver-Meal algorithm (see Axsäter [5]). In the same fashion, simple forecasting methods may be applied to complex demand patterns for the sake of simplifying the implementation and management of the forecasts. To understand how a method will respond to an environment it was not designed for, it is necessary to understand the entire process, from planning environment to measurement of results. As it may be difficult to predict how a system based on a required type of input will react to non suitable data, a model of the system may help to give insight into the system’s performance. Ashby [8] states in his law of requisite variety, that a regulatory system must have at least as much variety as its input to fully control the outcome; applied to the inventory control process, this effectively means that all aspects of a system must be modelled to give an accurate result. Inventory control systems consist of a three types of methods; forecasting, safety stock sizing and order timing and sizing [5]. Though there are many individual methods, only one method of each type may be employed in an inventory control system for a single stock-keeping unit.

The research objective of this study is to provide an increased understanding of the following research questions: (1) How may the process of inventory control be described unambiguously in a framework that allows for any combination of methods to be used, and (2) Is there any benefit to treating inventory control as a process, rather than as individual methods? This paper describes how complete inventory control systems can be modelled by using a framework that maps all relations from planning environment to performance measures. The unambiguosity means that a model designed for simulation purposes can be transferred to a real-world implementation without loss of functionality. It also means that single parts of inventory control systems can be modified, so that a better understanding of how a specific part contributes to the overall performance of the system may be gained. The structure of this paper is as follows: First, Section 2 describes the research approach. Thereafter, Section 3 integrates existing theory to describe a framework for designing inventory control models. Section 4 introduces empirical data from a company, of which the planning environment was used in Section 5 to develop a simulation model based on the framework. Section 6 describes the results of simulations of individual methods, and one simulation based on the framework. After that, Section 7 discusses the implications of the results, while Section 8 describes the conclusions that can be drawn from the study.

2. RESEARCH APPROACH

The design of the framework was based on a literature review of inventory control theory and on the authors’ experience of the area. Simulations were run for two reasons: One was to simulate inventory control systems for planning environments that were either following, or not following the assumptions of the simulated method; this
Design of a Framework for Inventory Control – Evaluation of Forecasting and Inventory Control Systems

tests whether the system performance changes when a foreign planning environment is introduced. The other reason to simulate was to see if the framework could be of practical use when employed to help a case company decide on a for their planning environment appropriate inventory control solution.

3. DESIGN OF AN INTEGRATED FRAMEWORK

The design of the framework is based on observing how inventory control methods operate, what input the methods require and what output they give. The underlying assumption for inventory control systems is that there for any given time \( t \), is an inventory level \( LL_t \), which is reduced by demand \( D \) and increased by replenishment \( R \). Another assumption is that time is divided into buckets (as described by Pidd [9]), for continuous systems buckets are infinitesimal, and that for each bucket the lowest inventory level, which is sufficient to evaluate the effects of inventory control, is governed by Formula 1. The relationship between these factors has been deduced from the rules that material requirements planning is built on [6].

\[
LL_t = LL_{t-1} + R_{t-1} - D_t \quad \text{subject to } LL_t \geq 0
\]  

(1)

where, \( LL_t \) = lowest inventory level at time \( t \), \( R_{t-1} \) = replenishment quantity occurring before \( t \) and \( D_t \) = demand during \( t \).

Formula 1 is the law that dictates how the transactions of any system placed in the framework will operate. It considers replenishment to occur between time buckets, meaning that it is sufficient to monitor the lowest inventory level for the law to reflect inventory transactions. Information such as service levels, inventory position and the highest stock level may be calculated from the lowest inventory level. The formula governs the inventory transactions of any inventory control system, and must be represented in any inventory control application. All other parts of an application may vary, either depending on the planning environment in which an inventory control system is used, or on the design of the system. Figure 2 shows the framework, which starts with the planning environment and ends with a measurement of the system’s performance.

Figure 2. Framework describing relations within inventory control systems.

The planning environment comprises the characteristics of all aspects that may affect the timing/sizing decisions [2]. For each time unit, the environment, which determines the distribution of demand, generates momentary demand that is passed on to a forecasting method, to an inventory control method and to actual transactions. The type of demand, which is dictated by the planning environment, tells whether a backlog can be implemented or not, and what function that may represent it [4]. Forecasting is affected by past demand information and the planning environment [1]. The former is used to do time series analysis, which is common practice in inventory control, while the latter may concern other input, such as information that may improve forecasting or data needed for causal forecasting. The environment may also tell of changes in the demand pattern, which may necessitate adjusting forecasting parameters.
or changing the forecasting method. It is necessary to consider the aggregation level of data, as longer term forecasts will have a low coefficient of variation, at the cost of losing forecast responsiveness [6]. Data from forecasting is necessary for inventory control methods (forecasted mean values), and for safety stock sizing (forecast variability) [3]. Safety stock sizing is a method which buffers against deviations from the expected mean value of the forecast [4]. The assumption of safety stock sizing is that all forecasts are correct estimations of the future mean value of demand; any deviations from the forecast are attributed as demand variability. This effectively means that an ill-performing forecast simply detects higher demand variability than a good forecast. The sizing is also affected by the planning environment, as the uncertain time determines the need for safety stock, as lead times also may have variability, and as the environment will determine to what extent customers accept shortages, or low service [1].

Inventory control methods rely on forecasts, on safety stock sizing and on the planning environment. The safety stock is used as a cushion to maintain service, while forecasts and data from the planning environment, which are ordering costs, holding costs and lead times, are used to determine when and/or how much to order [6]. The actual balancing of supply, which comes from the replenishment of the inventory, and demand, which is sales or lost sales, takes place as inventory transactions. Measuring these transactions gives an understanding of how well an inventory control system performs for the given planning environment [4].

4. EMPIRICAL DATA

Data was collected from a local timber yard, which did not use any policy to manage its inventories. Because of built-in ERP functionality for the reorder point method and for the periodic order quantity method, these methods could be deployed at low cost. The question was if the methods could cope with the demand for timber, which was presumed to have trend and seasonal components. Based on an analysis of sales data, the demand for timber was found to be seasonal, but with no trend component. This information was used to generate a demand function, based on the normal distribution, for the simulation model. The purpose of the demand function was to allow the simulation model to run for longer periods of time (5 years). Real demand, as well as simulated, is shown in Figure 3.

![Figure 3. Simulated demand compared to actual demand.](image)

Demand characteristics are shown in Table 1 and other parameters pertaining to the planning environment are shown in Table 2.

<table>
<thead>
<tr>
<th>Demand</th>
<th>Year</th>
<th>Month</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>56316</td>
<td>4693</td>
<td>156</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>710</td>
<td>205</td>
<td>37</td>
</tr>
<tr>
<td>( s(\sigma^2) )</td>
<td>504100</td>
<td>42025</td>
<td>1369</td>
</tr>
</tbody>
</table>

As transport costs were considered to be semi-fixed rather than variable, the reordering cost is valid for the reorder quantity used. Increasing the order quantity was not a cost-effective option. Stock out costs were not considered, as the consequences of stock outs are hard to measure; not only are sales lost, there is also the possibility of competitors winning the sale, and of losing customers, as they cannot find what they need.
Table 2. Environmental parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead time (days)</td>
<td>7</td>
</tr>
<tr>
<td>Order cost (SEK)</td>
<td>3200</td>
</tr>
<tr>
<td>Holding rate (%)</td>
<td>20%</td>
</tr>
<tr>
<td>Unit cost (SEK)</td>
<td>4.49</td>
</tr>
<tr>
<td>Order quantity (units)</td>
<td>585</td>
</tr>
<tr>
<td>Order interval (days)</td>
<td>45</td>
</tr>
</tbody>
</table>

Lead times are considered as fixed, as no information on delivery timeliness was available. The expected cost for non-seasonal demand, assuming that no safety stock is used, would be 25 659 SEK for both the systems, with an expected fill rate of 98% for the reorder point method, while the fill rate for the periodic order quantity method would be 95% (calculated using the loss function, as described by Axsäter [5].

5. DESIGN OF A SIMULATION MODEL

To test how the framework could be applied to a real-world scenario, a simulation model was constructed with the purpose of evaluating inventory control solutions that were considered for implementation by a local timber yard. To support the inventory control systems, some simply managed forecasting systems were chosen as alternatives that may fit the inventory control solution. The combination of forecasting methods and inventory control methods, put into the context of the framework is shown in Figure 4.

All methods were verified against theory individually by testing if the method implementations gave the values that theory would dictate. Because of the time bucket sizing (days), undershoot of the reorder point occurred for the reorder point method; to handle this condition the reorder point was increased by ½ days of forecasted demand as suggested by Mattsson [10], granting near-theoretical performance of the method. Though several forecast methods were considered, the actual choice of forecast for the case company’s demand pattern was based on an evaluation of the mean absolute deviation. The mean deviation was also calculated to see whether a forecasting method followed the mean of demand accurately. Comparative results are shown in Table 3. The seasonally adjusted moving average (see Mattsson and Jonsson [3]) was chosen as the preferred method as it proved to be nearly as accurate as Holt-Winters method (see Axsäter [1]), while not requiring as careful calibration of the forecast parameters. Forecasts were monthly, and predicted the demand for the following month. The forecast value was multiplied by 1.5 to reflect
an economic inventory cycle time of 45 days. This simplification was done to see how the system would react to systematic design errors in the application of the forecast.

<table>
<thead>
<tr>
<th>Table 3. Summary of forecast errors.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>MAD</td>
</tr>
<tr>
<td>BIAS (%)</td>
</tr>
</tbody>
</table>

6. Simulation runs

To investigate how seasonality may affect the performance of simple planning methods, three different demand patterns with increasing complexity were simulated. In the first, demand was constant and without variance, in the second, variance was added, while in the third, light seasonality (±20%) was introduced. Each simulation was measured 5 consecutive years (runs), with stable behaviour observed through all years. Table 4 shows the values used for demand. A lead time of one day was used, while the order quantity was 900 with the corresponding interval 15 days.

<table>
<thead>
<tr>
<th>Table 4. Demand pattern for the theoretical cases.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand</td>
</tr>
<tr>
<td>μ</td>
</tr>
<tr>
<td>σ</td>
</tr>
<tr>
<td>s (σ²)</td>
</tr>
</tbody>
</table>

The forecast was fixed to reflect the yearly average in all three cases, meaning that the framework was bypassed to ensure that only the effect of seasonality on the methods would be measured. Ignoring safety stock altogether, the measures used for the simulations were fill rate and inventory cycle service level. Figure 5 shows the results for the three demand types.

![Inventory cycle service level and Fill rate](image)

Figure 5. Cycle service levels and fill rates.

What can be seen is that inventory cycle service is higher for the periodic order quantity method, while the reorder point method has a higher degree of inventory cycles with stock-out. However, the fill rate shows that the reorder point method is the best at fulfilling demand. Together, the measures tell that the periodic order quantity method has fewer, but significantly greater stock-outs than the reorder point method. Seasonality affects the severity of the stock-outs for the periodic order quantity method, but not for the reorder point method, which manages the variations in demand by shortening the inventory cycle time. We may therefore conclude that seasonality can have a noticeable effect on methods that are not designed for such demand patterns.

In contrast with the above simulations, the one which used information from the case company was based on the framework, integrating forecasts and inventory control methods. The result of the simulation is shown in Figure 6, which is based on fill rate and the total yearly cost of the inventory policy. The periodic order quantity system shows a considerably worse fill rate than the reorder point system, though the safety stock is sized so that both methods should have the same fill rate if demand were non-seasonal.
Design of a Framework for Inventory Control – Evaluation of Forecasting and Inventory Control Systems

7. DISCUSSION

The simulations indicate that the fill rate of the periodic order quantity method suffers when a seasonal demand pattern is introduced, while the reorder point method can maintain the same fill rate as if no seasonality were present. This is a result of the nature of the two methods, where variability affecting the reorder point method will affect the time of ordering, while the periodic order quantity method, with fixed ordering times, cannot regulate order timing to prevent stock outs. Instead, it must let the inventory level take the full effect of any variability. Conversely, the effect of variability on the reorder point method is that the resulting order interval may not be economic. Uneconomic order quantities are not always a severe problem, as shown by Axsäter’s investigation of less-than-optimal order quantities.

When comparing the methods used in the timber yard simulation using the framework, we find that the reorder point method is superior both concerning the safety stock costs, and the fill rate. The extremely low fill rate (77%) demonstrated by the periodic order quantity shows how inventory control methods that are unsuitable for the planning environment can be affected. The use of a monthly forecast not representing the next inventory cycle may also have contributed to the low fill rate. The simulation based on the framework helped give insight into how the inventory control process would react to the planning environment. It showed that a large safety stock would be required if the periodic order quantity were to be used. If applied over multiple products, the framework can tell if consolidation using the sensitive periodic order quantity system is less costly than the insensitive reorder point system. Given that the periodic order quantity system has a 100% uncertain time [1], it may be used as a benchmark in simulations, as variability and problems caused by poor design of the process always will be reflected in the fill rate. This means that efforts to improve an inventory control process can be tested using this method.

In the future we are planning to enlarge our simulation experiments by incorporating different kind of demand types (continuous and discontinuous) as well as new methods used in forecasting and ordering. Recent research works have shown that autoregressive and particularly GARCH forecasting methods outperform others in situations, where demand is fluctuating widely and has clearly “life-cycle” pattern [11]. Similarly recent purchasing order method research has argued that not a single ordering method should be used (so basically it is not a question, which method is the best one, but which one is most suited to certain environment), but usually combination of different purchasing methods should be incorporated in ERP systems during the entire life-cycle of a product [12]. However, if volumes are low, then even economic order quantities/reorder point systems, and periodic order policies should be abandoned; lot for lot policy might produce best results in these situations [12]. Thus, much depends from the operations strategy (order or stock based system), and from the amount of time, which customers are willing to wait for a delivery to reach their facilities [13].

8. CONCLUSIONS

Treating inventory control and forecasting as separate activities, while not acknowledging how forecasting and its application may affect inventory control may lead to incorrect assessments of a system’s performance in a certain planning environment. Approaching inventory control as a process that starts with a planning environment and ends with a measurement of the system’s performance shows that all activities are related, and that the end result may be affected by the activities or by the way they are connected. In this paper a simulation model was built, to show how
the framework could be implemented, and simulations were run. It was shown, that using the framework to design an inventory control process enables different systems to be compared, and gives measures of how a complete inventory control process performs. Using the framework to design simulation models for testing a system for a planning environment will give a clearer view of how the system will react, than if forecasts and inventory control methods were evaluated alone. Describing the activities and the connections presented in the framework gives a complete system that may be implemented exactly as described. Documenting a system in this manner, enables transfer from simulation models to real inventory control applications or vice versa. Using simulation models to evaluate changes to the planning environment, or a change in the methods used is also easily done, due to the modular design of the framework.

REFERENCES

Optimizing Product Placement for Manual Picking in Warehouses

Uday Venkatadri and Sachin Kubasad

1Department of Industrial Engineering
Dalhousie University
Halifax, Nova Scotia, B3J 2X4, Canada

ABSTRACT

This paper looks at the problem of optimal product placement for a rectangular manual picking area in a warehouse, motivated by a practical industrial case in the food and beverage industry. In this paper, we compare the operational implication of three heuristic product placement strategies for two types of layouts: one with dead-end aisles and the other with open ended aisles. Our results show that open-ended aisles result in less travel distance, which agrees with intuition. We did not find appreciable differences in the performance of the three assignment heuristics; they all seem to be good. However, it is important to note that certain assumptions were made in this paper in obtaining these results.

1. INTRODUCTION

It is well known that improving the product placement in a warehouse can lead to gains in productivity in the picking operation. We consider the problem of optimal product placement problem for a rectangular warehouse in which picking is done manually by an order picking crew which uses a predetermined picking route. We also assume that the picking is done for one (usually large) order at a time.

This problem has practical importance; for existing warehouses, it is important to know the savings that can result from a combination of improved aisle design and product placement. For new warehouses, the question is how to lay out the aisles and place products.

Our involvement in this was motivated by a practical industrial case in the food and beverage industry. A section of a company’s distribution centre (DC) came under review. The company wanted to know if the product placement could be improved. From the aisle design point of view, the section in question at the DC had dead-end aisles. We believed that this was resulting in higher total travel for picking. We were interested in looking at the impact of changing the aisles so that each picking aisle was accessible from either end.

From an aisle design perspective, we look at both open-ended aisles and dead-end aisles.

From a product placement point of view, we identified three heuristics that are commonly used in rectangular warehouse sections:

- Place products in a zig-zag manner along the aisles in decreasing order of annual volume to match the S-Shaped heuristic picking route commonly employed in practice (N-S heuristic),
- Place products in decreasing order along aisles, starting with the aisle closest to the picking origin (N-N heuristic), and
- Place high volume products as closely as possible to the origin (nearest location heuristic).

We first discuss the main results in the literature dealing with product placement in manually picked warehouses. We will then report on the results of our study in which we evaluated combinations of product placement heuristics and aisle for the practical problem.

2. RELEVANT LITERATURE

Gu et al. [1] present an excellent review of papers on warehouse operations. The framework presented in the paper essentially consists of two large blocks: warehouse design and warehouse operations. The two blocks are shown to then interact with each other with a bi-directional link. In other words, warehouse operations are affected by its design and vice versa.
Hall [2] looks at three strategies for picking in a simple rectangular warehouse: traversal, midpoint, and largest gap. He then develops approximations for expected travel distance for each of these picking strategies. In all cases the author considers open-ended aisles. In the traversal strategy, once picking starts on an aisle, the entire aisle is traversed. In the midpoint strategy the aisle is broken into two halves and the halves are accessed from either end. It is used when an aisle contains less than one pick. The largest gap strategy is similar to the midpoint strategy, except that the order picker enters as far as the largest gap between two adjacent picks. For each of these strategies, the author develops formulas for expected pick lengths. One assumption in the paper is that the picking locations are randomly distributed, independently and uniformly through the warehouse. This assumption is a restriction for dedicated warehouses, including the one we were interested in.

Roodbergen et al. [3] present an eleven-component estimate for travel distance estimation in warehouse picking operations. Their estimate is based on warehouses with any number of rectangular blocks assuming that the S-shaped heuristic is used for routing. However, random storage is assumed in this paper. They then compare the performance of their travel distance estimate with simulation and report that the average error compared to simulation for a test set of 320 layout problems is 0.3% for the S-shape heuristic with a maximum error of 2.9%.

Another relevant article to this research is the one by Sadowsky and Ten Hompel [4], who use an analytical method to calculate travel distance. They assume that the access frequency distribution from the front of an aisle is known. For example, this distribution could follow the exponential distribution, such that the probability of picking along the aisle is exponentially distributed as a function of the distance. One assumption of this study is that all aisles have the same access frequency.

Based on these papers in the literature, we concluded that a new model for estimating travel distance was necessary in our case to take into account the following aspects in combination:

- Random storage is too restrictive as an assumption. In our problem, the storage locations are dedicated because of the choice of technology and a desire for simplicity.
- The frequency of accessing the aisles could be different since the assignment of products to aisle is a decision variable for the warehouse planner.

2. Layout Structure

As mentioned earlier, we look at two types of layouts in this paper – one with open-ended aisles (Figure 1) and the other with dead-end aisles (Figure 2). We assume that the racks are structured in a block with a depot located at the bottom left. All picking routes begin and end at the depot. We also assume that the aisle length is much larger than the rack width as well as the spacing between aisles, such as the one between Racks 2 and 3.

![Figure 1: Open-ended aisle configuration allowing picking access from both ends](image)
In the configuration shown in Figure 2, an aisle may only be accessed one way. Once picking is completed along an aisle, the picker must backtrack along the aisle before the next pick.

Apart from the configuration of the aisles, products need to be assigned to them. We consider dedicated storage in this paper. While it is possible to develop optimization models for product assignment, we used three simple heuristics. The N-S assignment is shown in Figure 3. In this assignment, the highest volume products are allocated along Rack 1 in the northerly direction followed by Rack 2 in the southerly direction and so on. In this and all other heuristics, we do not allow product segmentation. In other words, the assignment of a product is such that it cannot be assigned to two aisles. Let us consider the example in which four products have been assigned to Rack 1. The fifth product cannot be assigned to Rack 1 because the number of remaining slots in the aisle is smaller than what is required for product storage. The product is then assigned to Rack 2 in a southerly direction. If the sixth product can be fit in Rack 1, the assignment is made, if not, assignment continues on Rack 2.

![Figure 2: Dead-end aisle configuration allowing picking access from only one end](image)

The N-N heuristic is shown in Figure 4. With this heuristic, products are assigned in a northerly direction across all aisles. This procedure results in decreasing product demands along racks which is an assumption in Sadowsky and Ten Hompel [4]. However, the expected probability of accessing an rack decreases with the rack number.

![Figure 3: N-S product assignment heuristic](image)
The third and last heuristic considered in this paper is the nearest location heuristic (Figure 5). Here, products are first sorted in descending order of volume. The slots are sorted in ascending order of distance from the depot. The first product in the product list is assigned slots starting with the closest slot in the slot list where the assignment results in the allocation of product to continuous slots within a single rack. The product is dropped from the product list and the slots from the slot list and the procedure continues.

In the example illustrated in Figure 5, products P1, P2, and P3 are in decreasing order of volume. Therefore, P1 is first assigned to the slots indicated. P2 is then assigned to slots starting with the closest available slot, which is the third slot in Rack 1. P3 is assigned next and this process continues. If P2 could not have been fit on Rack 1 because of the number of slots remaining in the rack after assigning P1, the assignment would have started in the first slot of Rack 3 (where P3 is shown to start).
2. DISTANCE APPROXIMATIONS

In this section, we develop distance approximations for the two configurations shown in Figures 1 and 2.

2.1. OPEN-ENDED CONFIGURATION

For the open-ended aisle configuration, we assume that once the order picker enters the aisle, the entire aisle is traversed. Obviously, this is a simplifying heuristic which does not take into account backtracking from an aisle depending on which aisle is entered next and where the next pick occurs.

Let the probability of entering aisle $j$ be $P(j)$. Let $Q$ be the set of all products and $Q_j$ the set of products in aisle $j$. Let $D_i$ be the demand rate of product $i$. The aisle is entered if any of the products in the aisle is required. Therefore:

$$P(j) = \prod_{k \in Q_j} (1 - \sum_{i \in Q} D_i)$$ (1)
Figure 7: Probability tree of traversal path for the open-ended configuration

Figure 6 shows the critical locations reached in a complete picking path. For the twelve rack configuration shown in the figure, there are 13 locations numbered 0, 2, … 12 for convenience. Once the picker starts at the depot, he or she has two choices; if Rack 1 needs picking, the picker ends up in location 0c (where c denotes ‘complete’); else, the ending location is 1i. In the former case, the destination node of the picker is labeled 0c because aisle A has been picked at this point. We label the destination node of the picker in the latter case as 1i (i stands for ‘incomplete’) because at this point, racks 2 and 3 (along aisle B) are not yet picked. Obviously, after picking along rack 1, the picker may go back to the start location or move to 2i (when location 2 is reached in this manner, the aisle in front has not been picked and hence the location is labeled 2i). Location 2 may also be reached from location 1 after picking on racks 2 and 3 (along aisle B) and in this case, we label the destination node 2c (where c denotes ‘complete’). In this manner, it is possible to build the entire probability tree of the picking path (refer to Figure 7). In Figure 7, P(E) represents the probability of entering the corresponding aisle and P(NE) represents the probability of not entering the aisle (the probabilities are calculated from Equation 1). Also, the depot location S is shown on both the left and right sides of the tree for the purpose of convenience.

Since the distances are known for each location, it is possible to estimate the expected distance by reducing the tree. From S, two locations are possible, 1i and 0c. P(S-1i) is the probability of going from the depot (location S) to location 1i. The expected distance contribution along this path is P(S-1i)D_{S,1i}, where D_{S,1i} is the distance from S to 1i. From 1i, it is possible to go to 2c or 3i and from 0c, S or 2i are the only possible nodes. The expected distance for the entire picking path can be calculated by first finding the probability of entering a node in Figure 1 and multiplying it by the probability of each arc going out of the node and the distance along the arc.

2.1. DEAD-END CONFIGURATION

For the dead-end configuration, we calculate the expected depth of traversal into each aisle. This is given by:

\[ ED(j) = 2 \times \frac{\sum_{k \in Q_j} CD_{kj} D_k}{\sum_{k \in Q_j} D_k} \]  \hspace{1cm} (2)

Here, \( ED(j) \) is the expected travel distance along aisle \( j \). \( CD_{kj} \) is the distance to the centroid of product \( k \) along the aisle. As before, \( D_k \) is the demand rate of product \( k \) which is stored in aisle \( j \). We then assume that the picker travels along the cross aisle step by step and enters each picking aisle and returns to the cross aisle, as necessary.
The traversal path for the dead-end configuration is shown in Figure 8. In this configuration, there are seven critical locations numbered 1, 2, 3, …, 7 for convenience. The aisles have been renamed A, B, C, …, G. The order picker begins a route at the depot and reaches location 1. At this point, there are two possibilities. If there are items to be picked in aisle A, the traversal distance along aisle A (in and out) is calculated using the formula in Equation 2. If not, the picker simply continues on to location 2. If aisle A is traversed, the picker may either go back to the depot or continue on to location 2. This depends on whether there are items to pick in aisles B, C, D, …, G. Equation 1 is used to calculate the probability of entering any one of these aisles. From this basic equation, the probability of entering any of these aisles may also be calculated. The probability tree for dead-end travel is shown in Figure 9. The expected travel distance is computed by multiplying the node probabilities by the arc probability and distance. This computation is very similar to that of the open-ended case, the only difference being the probability tree itself.

![Figure 8: Critical locations in picking path for the dead-end configuration](image1)

![Figure 9: Probability tree of traversal path for the dead-end configuration](image2)
3. RESULTS

We used the distance approximation calculations developed in this paper to look at placement strategies for an anonymous industrial partner’s distribution centre. There were 854 products for assignment in a rectangular section of the DC with the same number of racks and aisles as shown in Figure 6. However, the depth of the racks and aisles were different - there were a total of 933 storage locations in the example with two pallet storage in each location. The demand of products varied significantly. With slot locations proportional to annual demand, the number of slots required for each product varied from 0.3 to 54.7. We assumed that the number of slots required could be fractional, allowing more than one product to share a slot location. In the configuration, the aisle widths were 2.74m with the racks themselves being 51.5m. Picking always takes place at the floor level locations, even though storage is 3 levels high. Table 1 below shows the distance approximation calculation results.

Table 1: Distance Approximations (metres)

<table>
<thead>
<tr>
<th>Layout Configuration</th>
<th>Assignment heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N-S</td>
</tr>
<tr>
<td>Open-ended</td>
<td>459.99</td>
</tr>
<tr>
<td>Dead-end</td>
<td>639.08</td>
</tr>
<tr>
<td>Increase</td>
<td>38.93%</td>
</tr>
</tbody>
</table>

As can be seen from the table, the nearest-location performs best overall when the open-ended configuration is used. However, there is very little difference between this and the other heuristics. In fact, the worst heuristic for this case is the N-S heuristic, whose performance is only 1.93% worse than the nearest-location heuristic. Our speculation is that the nearest-location heuristic keeps the location of frequently picked products as much as possible to the bottom and to the left (i.e., towards the depot). Therefore, it might result in the least expected distance. Our assumption is that the entire aisle is travelled when entered in the case of the open-ended configuration. This is obviously an oversimplification. We are in the process of examining what happens if the aisle can be broken up into two or more partitions to improve the distance estimate.

From Table 1, it can also be seen that the dead-end configuration results in much higher expected travel than the open-ended configuration. For example, in the N-N case, the performance degradation is 52.38%. For the dead-end configuration, the nearest location heuristic provided the least expected travel. We believe again that this heuristic keeps frequently visited products as close as possible to the depot, possibly the main reason why it is the best. The differences in performance between the three heuristics are appreciably more than in the open-ended case. The main benefit to the industrial partner from this study was that we were able to show them that it is better to use the open-ended configuration, though it requires slightly more space. By opening up the aisles, we expect that they will be able to improve their picking efficiency by at least 36%.

4. CONCLUSIONS AND FURTHER STUDY

Optimizing product placement has very important implications for picking operations. In this paper, we tried to build a costing model to look at the impact of three product placement heuristics on two different aisle configurations. It is clear that the open-ended aisle configuration results in less expected travel. However, our research is still in progress. The calculation of expected distance needs to be refined in many ways, especially for the open-ended configuration. We assumed in the open configuration model that once an aisle is accessed, it is traversed completely. This is not true. Depending on which products need to be accessed in the next aisle, it may be better to backtrack. We are in the process of modeling this further. We are also trying to build a simulation model to gain more insights and improve our expected distance calculations. Finally, we are interested in looking at how cross aisles may help improve picking efficiency.
ACKNOWLEDGEMENTS

The support made available to Uday Venkatadri through the National Science and Engineering Research Council of Canada’s Discovery and Equipment grants programs is greatly appreciated. Thanks are also due to an anonymous industrial partner who provided the data for the analysis done in this paper.

REFERENCES


An exact approach to minimize the sum of maximum earliness and tardiness in single machine scheduling problem with sequence-dependent setup time

Nooshin Nekoiemehr* & Ghasem Moslehi

Department of Industrial and Systems Engineering,

Isfahan University of Technology,

Isfahan, 84156-83111, Iran.

Abstract

This paper considers the problem of scheduling a given number of jobs on a single machine to minimize the sum of maximum earliness and tardiness when sequence-dependent setup times exist \( (1|\text{ST}_{sd}|\text{ET}_{\text{max}}) \). This scheduling problem in view of trying to minimize the values of earliness and tardiness, can be applied in different production systems such as just in time (JIT) systems. Though considering sequence-dependent setup times increase the complexity of the problem significantly, is often encountered in different production systems. In this paper an optimal branch and bound algorithm is developed including the implementation of lower and upper bounding procedures, and dominance rule. For solving problems containing large number of jobs a polynomial time-bounded heuristic algorithm is also proposed. The proposed algorithms were empirically evaluated on 3840 problems of various sizes, ranging from 8 to 25 jobs, and different parameters. Computational experiments demonstrate the effectiveness of the bounding rules in achieving optimal solutions in more than 97% percent of the instances.

1. Introduction

In this paper, we address the problem of finding an optimal solution of \( n \) jobs with sequence-dependent setup times to minimize the sum of maximum earliness and tardiness on a single machine; the problem is shown as \( (1|\text{ST}_{sd}|\text{ET}_{\text{max}}) \). Interest in scheduling problems that treat setup times (costs) as separate began in the mid-1960s [1]. The importance of setup times has been investigated in several studies. Panwalkar [2] found about three quarters of the managers in his survey reporting that at least some operations they scheduled required sequence-dependent setup times, while approximately 15% reporting all operations required sequence-dependent setup times [1]. Scheduling problems with sequence-dependent setup times are often encountered in many production systems and industrial applications including color procedure in the plastic industry [3], wafer testing in semiconductor manufacturing [4], and in photocopiers and laser printers [5]. Allahverdi et al. [1] surveyed the literature on separate setup times involving static, dynamic, deterministic, and stochastic problems for all shop environments including single-machine, parallel machines, flow shops, job shops, open shops, and others.

Wisner and Siferd [6] reported that meeting target delivery dates was the most important scheduling objective and only 58% of industrial schedulers managed to meet delivery dates. As declared by Luh et al. [7], “the understanding of sequence-dependent setups, however, is still believed to be far from being complete, especially of scheduling to meet due dates”. Meeting due dates while also considering both tardiness and earliness is important. A number of authors have addressed problems involving due dates and sequence-dependent setup times in order to minimize sum of earliness and tardiness. Baker and Scudder [8] provide a survey of sequencing problem with earliness and tardiness penalties while Kanet and Sridharan [9] provide a review of scheduling with inserted
In the remainder of the paper, \( \sigma' \) refers to the set of jobs that are not scheduled. Job \( \sigma(j) \) is defined as a partial schedule and \( \sigma' \) as a set of jobs that are not scheduled. Job \( \sigma(j) \) itself, whereas job \( \sigma(j) \) (with brackets) refers to the job \( j \) itself. In the remainder of the paper, \( \sigma \) is assumed that all jobs are available from time zero and the machine has no predefined initial setup setting whatever the first task is. In addition, The triangle inequality is satisfied; in other words, among the setup times jobs \( i,j,k \), the inequality \( s_{ij} + s_{jk} \geq s_{ik} \) is applied. In a single machine scheduling, earliness \( (E_j) \) and tardiness \( (T_j) \) of job \( j \), maximum earliness \( (E_{max}) \), maximum tardiness \( (T_{max}) \), and the sum of maximum earliness and tardiness \( (ET_{max}) \) in each sequence are obtained as follows:

\[
E_j = \max(0,d_j - C_j), \quad j = 1,...,n
\]

(1)

\[
T_j = \max(0,C_j - d_j), \quad j = 1,...,n
\]

(2)

\[
C_j = \begin{cases} 
0 & j = 0, \\
\sum_{k=1}^{j} (s_{k-1} + p_k) & j \neq 0.
\end{cases}
\]

(3)

\[
E_{max} = \max_{1 \leq j \leq n} \{E_j\},
\]

(4)

\[
T_{max} = \max_{1 \leq j \leq n} \{T_j\},
\]

(5)

\[
ET_{max} = E_{max} + T_{max}.
\]

(6)

Although \( ET_{max} \) is a non-regular measure and it may be possible to reduce the objective by inserting idle time into a schedule, we assume that no idle insert is allowed as already mentioned in our assumptions above.

### 3. Branch-and-Bound Algorithm
To find the optimal sequence, we propose a branch-and-bound implicit enumeration approach for the problem of \( |ST_{ad}|ET_{max} \), the proposed branch-and-bound algorithm involves back-tracking rule as an branching procedure and each sequence is constructed by starting from the beginning and working forward so a node at the \( K \) level in the B&B tree corresponds to a partial sequence for the first \( K \) jobs in a sequence, \( \sigma = J_1 \ldots J_k \).

### 3.1. LOWER BOUND

In order to develop a lower bound for the problem of \( |ST_{ad}|ET_{max} \), a lower bound is initially described in lemma 1 for the problem of minimizing the sum of maximum earliness and tardiness without setup times (\( |J|\}) \) ET_{max}). Then, based on the procedure presented by Schaller, lower and upper bounds are proposed on the completion times of unscheduled jobs by means of setup times \([17,22]\). Next, these bounds are employed in a mathematical program whose optimal solution provides a lower bound for the problem of \( |ST_{ad}|ET_{max} \).

Let \( \text{d}_{\text{EDD}}[j] \) equal the due date of the \( j \)th job. If the jobs are sorted according to the earliest due date (\( \text{d}_{\text{EDD}}[j] \leq \text{d}_{\text{EDD}}[k] \) if \( j < k \)) and if the completion times of the job in each position of a sequence are known, then a lower bound is obtained on the sum of maximum earliness and tardiness without setup times from the following theorem.

**Lemma 1.** If the completion time of the jobs in each position of a sequence is known, \( LB \) is the lower bound for the problem \( |J|\}) \) \( ET_{max} \). Where, \( C_{[ij]} \) is the completion time of the job in \( j \)th position of the sequence.

\[
LB = \max \left( \max \left( C_{[ij]} - \text{d}_{\text{EDD}}[j],0 \right) \right) + \max \left( \max \left( \text{d}_{\text{EDD}}[j] - C_{[ij]},0 \right) \right)
\]

**Proof.** Consider a schedule \( S \) in which the job at the sequence position \( j \) has a due date greater than the due date of the job in the sequence position of \( k \), where \( j < k \), if \( d_0[j] \) and \( d_0[k] \), respectively, are the due dates of the jobs in the sequence positions \( j \) and \( k \), then \( d_0[j] < d_0[k] \). Now consider another schedule \( S' \) obtained by interchanging due dates of these two jobs in these positions. In other words, \( S' \) is obtained by assigning the due date of the job in position \( k \) to that in position \( j \) and vice versa. Therefore, the completion times for the jobs in all positions and their due dates will be the same except due dates of jobs in two positions of \( j \) and \( k \) in both schedules. If \( d'_{[i]} \) represents the due date of the job in the sequence position \( i \) in schedule \( S' \), then \( d'_{[i]} = d_0[i] \) if \( i \neq j \) and \( i \neq k \), \( d'_{[j]} = d_0[k] \), and \( d'_{[k]} = d_0[j] \). The tardiness and earliness of the jobs in schedules \( S \) and \( S' \) will be defined as follows. \( T_{[i]} = \max(c_{[i]} - d_0[i],0) \), \( E_{[i]} = \max(c_{[i]} - d_0[i],0) \), \( T_{[j]} = \max(c_{[j]} - d_0[j],0) \), \( E_{[j]} = \max(c_{[j]} - d_0[j],0) \), \( E_{[k]} = \max(d_0[k] - c_{[k]},0) \), and \( E_{[k]} = \max(d_0[k] - c_{[k]},0) \). Since \( T_{[m]} = T'_{[m]} \), \( E_{[m]} = E'_{[m]} \) if \( m \neq j \) and \( \text{min} \) earliness and tardiness of jobs in the two positions \( j \) and \( k \) are considered. Considering that the positions of jobs in the two schedules are the same, the completion times will not be different and, according to the relations between the due dates of jobs, we will have

\[
T'_{[j]} = \max(c_{[j]} - d'_{[j]},0) = \max(c_{[j]} - d_0[i],0) < T_{[j]} = \max(c_{[j]} - d_0[j],0)
\]

\[
T'_{[k]} = \max(c_{[k]} - d'_{[k]},0) = \max(c_{[k]} - d_0[k],0) < T_{[k]} = \max(c_{[k]} - d_0[j],0)
\]

\[
E'_{[j]} = \max(d'_{[j]} - c_{[j]},0) = \max(d_0[j] - c_{[j]},0) < E_{[j]} = \max(d_0[j] - c_{[j]},0)
\]

\[
E'_{[k]} = \max(d'_{[k]} - c_{[k]},0) = \max(d_0[k] - c_{[k]},0) < E_{[k]} = \max(d_0[k] - c_{[k]},0)
\]

Therefore, Equations (8) to (11) will yield \( \max(T'_{[j]}, T'_{[k]}) \leq T_{[i]} \) and \( \max(E'_{[j]}, E'_{[k]}) \leq E_{[i]} \). We will also have \( \max(T'_{[j]}, T'_{[k]}) + \max(E'_{[j]}, E'_{[k]}) \leq \max(T_{[j]}, T_{[k]}) + \max(E_{[j]}, E_{[k]}) \). Consequently, it may be concluded that the sum of maximum earliness and tardiness in schedule \( S' \) is not larger than that of schedule \( S \).

#### 3.1.1. COMPLETION TIME BOUNDS

As we do not know the completion times of unscheduled jobs in the branch-and-bound algorithm, as required by lemma 1, we will use setup times to determine lower and upper bounds on the completion times of unscheduled jobs. These bounds are then used in a mathematical program whose optimal solution provides a lower bound for our problem \( |ST_{ad}|ET_{max} \).

To compute lower and upper bound for completion times of unscheduled jobs, first allowing \( pS_{PT_{[j]}}(\sigma') \) and \( pL_{PT_{[j]}}(\sigma') \) to equal the processing time of the job in position \( j \), respectively, if the jobs in \( \sigma' \) are sorted according to the shortest processing time, and longest processing time. If we pick the minimum amount of each job’s setup time, \( S_j = \min(S_{ij}) \) \( \forall i \), then \( S_{SST_{[ij]}}(\sigma') \) equals the selected setup time for the job in position \( j \), once the selected setup times are
sequenced according to the shortest setup time \( S_{SST[i]}(\sigma') \leq S_{SST[j]}(\sigma') \) if \( a < b \). If we pick the maximum amount of each job’s setup time, \( S_j = \max(S_i) \forall i \), then \( S_{LST[i]}(\sigma') \) equals the selected setup time for the job in position \( j \). Once the selected setup times are sequenced according to the longest setup time \( S_{LST[i]}(\sigma') \geq S_{LST[j]}(\sigma') \) if \( a < b \). So, \( LC[j] \) and \( UC[j] \) will be respectively the lower and upper bound of the completion time for the job in position \( j \).

\[
\begin{align*}
LC(j) &= \sum_{k=1}^{j} p_{LPT(k)} + s_{LST[j]} \quad (12) \\
UC(j) &= \sum_{k=1}^{j} p_{LPT(k)} + s_{LST[j]} \quad (13)
\end{align*}
\]

**Lemma 2.** \( LC[j] \) and \( UC[j] \) are respectively the lower and upper bound on the completion time of a job in position \( j \) of any sequence of unscheduled jobs.

**Proof.** Although the proposed lower and upper bound on completion times differs from that proposed by Schaller, in that we pick the minimum and maximum setup times of each job and also that there is no setup family for the jobs, the proof for our proposed lower and upper bound on completion times can be derived along the same lines as Schaller’s [19] since both have identical basic structures.

### 3.1.2. Lower Bound of \( l|ST_{sd}|ET_{max} \)

To find the lower bound for the problem of \( l|ST_{sd}|ET_{max} \), the following mathematical program is proposed.

**Program \( ET_{maxEDD} \).**

\[
\begin{align*}
\min & \quad z = \max_{j} (\max(c_{i,j} - d_{EEDD(j)}, 0)) + \max_{j} (\max(d_{EEDD(j)} - c_{i,j}, 0)) \\
\text{subject to} & \quad c_{i,j} \geq LC_{i,j} \quad (15) \\
& \quad c_{i,j} \leq UC_{i,j} \quad (16)
\end{align*}
\]

It should be noted that the decision variables in this program are the completion times of jobs in each position of a schedule. The objective value was shown above is the lower bound for the problem of \( l|\ ) ET_{max} \) and the completion times are restricted by their lower and upper bounds which were calculated using setup times.

In the problem \( l|ST_{sd}|ET_{max} \), let \( \delta' \) be the optimal sequence of jobs in \( \sigma' \) and \( C'_{i,j}(\delta') \) be the optimal completion time of the jobs in the positions of the sequence \( \delta' \). If we show that the optimal completion times of the jobs in \( l|ST_{sd}|ET_{max} \) are feasible for \( ET_{maxEDD} \), then the optimal value gained from \( ET_{maxEDD} \) will be a lower bound for the problem \( l|ST_{sd}|ET_{max} \).

**Lemma 3.** \( C'_{i,j}(\delta') \) in the problem \( l|ST_{sd}|ET_{max} \) are feasible for \( ET_{maxEDD} \).

**Proof.** According to lemma 2, for any sequence of unscheduled jobs (\( \delta \)), \( C_{i,j}(\delta) \geq LC_{i,j} \) and \( UC_{i,j}(\delta) \). Therefore, \( C_{i,j}(\delta') \geq LC_{i,j} \) and \( C_{i,j}(\delta') \leq UC_{i,j} \). The constraints imposed by \( ET_{maxEDD} \) also restrict completion times between \( LC_{i,j} \) and \( UC_{i,j} \). So, the optimal completion times of the jobs in \( l|ST_{sd}|ET_{max} \) will be feasible for \( ET_{maxEDD} \) and the optimal solution of \( ET_{maxEDD} \) will be the lower bound for \( l|ST_{sd}|ET_{max} \). ■

The program \( ET_{maxEDD} \) explores the completion times of jobs in the positions of a sequence independently. Therefore obviously the algorithm for obtaining the optimal solution of \( ET_{maxEDD} \) as follows:

**Step 0.** Set \( k = n \), \( C_{i}| = LC_{i} | j = 1, \ldots, n \). Enter step 1.

**Step 1.** If \( d_{EEDD[k]} \leq C_{k} \), go to step 3; otherwise, go to step 2.

**Step 2.** If \( d_{EEDD[k]} \leq UC_{k} \), let \( C_{k} = d_{EEDD[k]} \) and go to step 3; otherwise, \( C_{k} = UC_{k} \) and go to

**Step 3.** Set \( k = k - 1 \). If \( k > 1 \), go to step 2; otherwise, STOP.
3.2. DOMINANCE RULES

In this section, we present a dominance rule. Good dominance rules could lead to substantial improvements in algorithmic performance. Consider \( u \), as the last scheduled job in \( \sigma \), and \( \sigma' \) as a set of unscheduled jobs.

Assume that we obtain a modified processing time for the unscheduled jobs (\( \sigma' \)) by adding each job’s minimum setup time to its actual processing time. Then, using the modified processing times, these jobs will be sorted in the reverse MST order. If the maximum earliness of this sequence is less than the maximum earliness of the partial sequence \( \sigma \), this condition will be named **worst maximum earliness**.

**Lemma 4.** If the condition **worst maximum earliness** exists, there is an optimal sequence in which job \( i \in \sigma' \) will be scheduled in the first possible position of the schedule based on the following assumptions. See Fig (1).

\[
\begin{align*}
    s_{u,j} + s_{j} - d_{j} & \geq s_{u,j} + s_{i,j} - d_{j} \quad \forall j \in \sigma' \quad (17) \\
    s_{u,j} + s_{j} + \min_{k \neq i}(s_{k,j}) & \geq s_{u,i} + s_{i,j} + \max_{k \neq i}(s_{j,k}) \quad k \neq i, j \quad (18)
\end{align*}
\]

**Proof.** It can easily be shown that sum of maximum earliness and tardiness in schedule \( \sigma_0 \) is less than \( \sigma_1 \), by considering assumptions (17) and (18), therefore \( \sigma_1 \) is dominated.

3.3. UPPPER BOUND

We propose a heuristic algorithm that consists of two phases. In the first phase, an initial sequence is created based on a greedy procedure. Phase two, then, attempts to improve this initial sequence using two improvement procedures such as pairwise exchange of adjacent jobs and insertion algorithms.

3.3.1 CREATING AN INITIAL SEQUENCE

An initial sequence is created by a greedy algorithm. In greedy algorithm the set of candidate jobs consists of all those that are yet to be sequenced. The next job to be sequenced is selected using an appropriate greedy cost function. The cost function presented here considers slack time and will be different for tardy or early jobs. The construction steps are as follows. Assume \( S \) is the partial sequence and \( S' \) is the set of unscheduled jobs. Let \( u \) be the last scheduled job.

**Step 1.** Set \( S = \emptyset, S' = n, t = 0 \).

**Step 2.** Evaluate the slack time \( (d_{i} - p_{i} - S_{u,k} - t) \forall k \in S' \) of candidate jobs, jobs in set \( S' \).

**Step 3.** If the slack time is positive, which means an early job, we use function \( COST_{f} = \frac{((d_{k} - p_{k} - S_{u,k} - t))}{(p_{k} + S_{u,k})} \) for calculating the scheduling cost of the job. If the slack time is negative, which means a tardy job, we use function \( COST_{f} = \frac{((d_{k} - p_{k} - S_{u,k} - t))}{(p_{k} + S_{u,k})} \) for calculating the scheduling cost of the job. Then, select the job with the least amount of cost and put it in \( S \) at time \( t \) and update following parameters \( t = t + p_{k} + S_{u,k}, n \leftarrow n - 1, S' \leftarrow S' - \{k\} \).

**Step 4.** If \( n = 0 \), stop; otherwise, go to Step 2.

3.3.2. IMPROVEMENT PROCEDURES

The first procedure attempts to improve a sequence by exchanging the positions of two adjacent jobs. If the maximum earliness and tardiness of a sequence obtained by exchanging positions is less than that of the previous
sequence, then the exchange is retained; otherwise, the exchange is reversed. This procedure will be continued up to the last scheduled job. The second procedure tries to improve the sequence by inserting the job with the maximum tardiness in the first position of the sequence, then to the second, and the process continues until the job is placed in the position immediately preceding its original position. Then, the sequence with the least objective is retained. Then, the sequence obtained from the previous procedure is improved by inserting the job with maximum earliness from the last position in the sequence, then to the last but one, and the process continues until the job is placed in the position immediately following its original position. Then, the sequence with the least objective is retained. The proposed heuristic algorithm can be implemented in polynomial-bounded time.

4. Computational experiments

This section presents the results of the empirical experiments designed to evaluate the effectiveness of the proposed algorithm. The algorithm was implemented in Microsoft Visual C++ 6.0 and ran on a PC with a 3.4 GHz Pentium IV CPU and 512 M RAM. The proposed algorithm was evaluated using problem instances of various sizes and based on control parameters given in Table 1. Each problem set consists of 10 problems. The processing times of jobs were random integers from a uniform distribution over the range \([10,100]\) based on methods proposed in \([19, 20, 21]\). We will use a parameter for the setup proportion factor (SPF, the proportion of the range of setup times to processing time) which was originally due to Lou \([21]\). This factor decides the range of setup time distributions according to the range of processing time distributions. So setup times were created from \([10, spf \times P_{\text{max}}]\), where \(P_{\text{max}}\) is the maximum processing time of jobs. Generally, there are two control parameters to generate due dates for instances: due date range (\(R\)) and tardiness factor (\(\tau\)). Due dates for jobs were also randomly generated over the range \([\max(T(1-\tau-R/2),0),T(1-\tau+R/2)]\), where \(T = \left(\sum_j p_j + \text{average}(S)\right)\) \([20]\).

<table>
<thead>
<tr>
<th>size</th>
<th>(R)</th>
<th>(\tau)</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>{8,10,12,15,20,25}</td>
<td>{0.2,0.4,0.6,0.8}</td>
<td>{0.2,0.4,0.6,0.8}</td>
<td>{0.2,0.6,1.1,1.4}</td>
</tr>
</tbody>
</table>

Sixteen different series of problems were generated by combining the two parameters \(R\) and \(\tau\). The total number of instances by parameter combinations will be 3840 (4×4×4×6×10).

4.1. Efficiency of the proposed algorithm

The measure of efficiency is the CPU time required to produce the solution. The maximum CPU time of each instance was set at 3600 seconds. The results of applying the B&B algorithm to all random instances show that all the problems with 8 to 15 jobs were optimally solved. The percentage of instances solved decreased with increasing job size. In fact, our algorithm was able to solve 93.56% of 20-job instances and 90.21% of 25-job instances. The average CPU time required for solving all instances in sizes 8,10,12,15,20,25 are respectively 0.02, 0.85, 43.64, 582.52, 1437.24, 1624.10. Increasing the problem size and setup time amounts increased CPU time, gradually reducing the number of optimally solved instances. Judging from the capability of the proposed branch-and-bound algorithm in optimally solving 97.5% of the instances with a size range of 8 to 25 jobs, it may be claimed that the proposed procedures in lower and upper bounds and dominance rule enjoys a high level of efficiency. Results in Table 2 present the average CPU time and also average percent of unexplored nodes caused by both lower bound and dominance rule in B&B tree for sixteen series of instances. By fathoming each node all its children in search tree will also be fathomed, calculating the number of the children on each node based on its level in the search tree regarded as an unexplored node. Of course the percent of unexplored nodes in different series shows the effectiveness of proposed lower bound and dominance rule.

<table>
<thead>
<tr>
<th>size</th>
<th>(R)</th>
<th>Average time (s)</th>
<th>Unexplored nodes</th>
<th>Average time (s)</th>
<th>Unexplored nodes</th>
<th>Average time (s)</th>
<th>Unexplored nodes</th>
<th>Average time (s)</th>
<th>Unexplored nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.2</td>
<td>0.04</td>
<td>73.47</td>
<td>0.04</td>
<td>60.71</td>
<td>0.04</td>
<td>64.24</td>
<td>0.04</td>
<td>61.98</td>
</tr>
<tr>
<td>10</td>
<td>1.74</td>
<td>1.87</td>
<td>78.33</td>
<td>1.96</td>
<td>78.18</td>
<td>1.75</td>
<td>81.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In order to calculate the efficiency of the heuristic method, the average of relative deviation (ARD) from the optimum value is considered that calculated as follows:

\[
ARD = \frac{1}{n} \sum_{j=1}^{n} \frac{H_j - OPT_j}{OPT_j} \times 100
\]  

(19)

Where, \(H_j\) and \(OPT_j\) are objective function values of the heuristic and the branch-and-bound methods, respectively. The results in Table 3 actually show that the average percentage of deviation in the heuristic algorithm was less than 5% in all problem sets but the average deviation from the optimal solution was about 2.78%, which confirms its efficiency.

<table>
<thead>
<tr>
<th>(\tau)</th>
<th>(0.2)</th>
<th>(0.4)</th>
<th>(0.6)</th>
<th>(0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.44</td>
<td>0.87</td>
<td>0.80</td>
<td>1.37</td>
</tr>
<tr>
<td>10</td>
<td>1.53</td>
<td>2.42</td>
<td>2.49</td>
<td>2.36</td>
</tr>
<tr>
<td>12</td>
<td>2.00</td>
<td>2.83</td>
<td>2.19</td>
<td>2.86</td>
</tr>
<tr>
<td>15</td>
<td>3.09</td>
<td>3.05</td>
<td>2.73</td>
<td>3.23</td>
</tr>
<tr>
<td>20</td>
<td>3.22</td>
<td>3.46</td>
<td>3.46</td>
<td>3.28</td>
</tr>
<tr>
<td>25</td>
<td>3.49</td>
<td>3.64</td>
<td>4.14</td>
<td>4.00</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper addressed the single machine scheduling problem with sequence-dependent setup times to minimize the sum of maximum earliness and tardiness. A branch-and-bound algorithm was proposed including efficient lower and upper bounds and a dominance rule. The upper bound here consisted of two main phases. First, creating an initial sequence based on a greedy procedure including proper cost functions and, second, an improving process for the initial sequence obtained that includes pairwise and insertion procedures. The lower bound was derived by determining bounds on completion times of unscheduled jobs and solving a simple mathematical program. Evaluating the efficiency of the proposed algorithm shows that the proposed algorithm was capable of optimally
solving 8-job to 15-job instances with efficiency of 100% as well as 20-job and 25-job instances with efficiencies of 93.56% and 90.21%, respectively.

REFERENCES

Least Angle Regression Using Orthogonal Basis of Input Variables

Dongil Kim, Jibin Hwang, Sung-Shick Kim, and Jun-Geol Baek*.

Graduate School of Information Management and Security
Korea University
Seoul, 136-713, Republic of Korea

ABSTRACT

In the paper, we propose a new variable selection algorithm called Adjusted LARS, based on LARS(Least Angle Regression). Variable selection is mainly used to make prediction model. A main cause to prevent prediction model from having accuracy, is multi-collinearity. Original LARS calculates correlation between input variables and residual to decide the priority of input variables. The main idea of the Adjusted LARS is finding an orthogonal basis of input variables by QR factorization and calculating correlation between basis and residual. The Adjusted LARS is expected to exclude effect of multi-collinearity and improve rate of correctly-selected variables. Experiment results indicate improvement of the performance of the Adjusted LARS compared to the original LARS.

1. INTRODUCTION.

Variable Selection is one of the basic topics in data mining. It determines input variable that will be used in prediction modeling. Variable selection is used in mathematical application like supervised or unsupervised machine learning, regression, time series prediction, two-class or multi-class classification. And it is also used in text processing of internet documents, gene expression array analysis, and combinatorial chemistry [1].

Mathematical definition of variable selection is as follow. There is output vector \( Y \) and Set \( N \) of input vectors \( \{X_1, \cdots, X_p\} \) which are expected to having linear relation with \( Y \). And arbitrary subset \( S \) of \( N \) is set of selected input variable. \( S \) made a linear space and \( \bar{Y} \) is vector created by projecting \( Y \) on the linear space. And residual \( R \) is Euclidian distance and defined as. \( (1) \)

\[
R = \sum_{i \in S} (Y_i - \bar{Y})^2
\]

Definition of variable selection as using \( \bar{Y} \) and \( R \), it is finding \( S \) which make minimum \( R \) and minimum subset size simultaneously.

Variable selection is NP-hard problem [2]. LARS is one of many heuristic algorithm for variable selection. It has fast computational speed, and its accuracy is better than SRF(Stepwise Regression Forward) which is widely used algorithm [3]. The algorithm proposed in the paper base on LARS, and we develop it for more correct variable selection.

Main issues of variable selection are as follows. It must help understand how input variables effect on output, reduce measure data and data storage requirement, reduce training time and using time, improve accuracy of model. The paper focuses on reducing measure data and improving accuracy of model.

In session 2, we explain disturbing elements of Variable selection. And in session 3, we review existing research such as stepwise regression and LARS. In section 4 we explain more details of these issues and offers solutions to resolve the problem. In session 5, we perform simulations and analyze the results, to verify the algorithm SRF, LARS, and to compare the algorithm proposed in the paper.

* Corresponding author: Tel.: (82) 2-3290-3396; Fax: (82) 2-929-5888; E-mail: jungeol@korea.ac.kr
2. ELEMENT OF THE INTERFERENCE

In variable selection there are two main causes which prevent prediction accuracy of algorithm. The one is over-fitting, the other is multi-collinearity.

If some algorithm chooses specialized variables for sampling data we call that variable selection is over-fitted. When we make some model, we use sampling data and we choice the model with some input variable that fit for sample data. But there is no guarantee that selected variables fit for population. Over-fitting occurs when statistical model has too many parameters. Especially, if sampling data is biased or includes singular value, statistical model differ with real model because of over-fitting. [3]

Multi-collinearity is divided into exact multi-collinearity and near multi-collinearity. Exact multi-collinearity means that input variable \( X \) is linearly dependent on some other input variable. Near multi-collinearity means that input variable \( X \) is not expressed by other input variable' linear combination, but expressed as follow (1) \( C \) is arbitrary vector which is not null-vector. [4]

\[
XC = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \approx 0
\]

(1)

Exact multi-collinearity is not an important problem in variable selection. Because it can be detected and removed easily by LU decomposition and singular value decomposition. But near multi-collinearity is not solved exactly. If input variables with near multi-collinearity exist in set of selected variable, the prediction model which consisted of that set is unstable [3]. And if we exclude input variables with near multi-collinearity collectively, we can exclude important variables that effect on output in combination with each other [1].

3. REVIEW EXISTING RESEARCH

Various heuristic algorithms for variable selection have been researched. Sometimes their prediction accuracy is prevented by over-fitting problem and near multi-collinearity. We handle improvement in that problems. Efromyon proposed stepwise regression which is widely-used heuristic algorithm [5]. LARS improves prediction accuracy of stepwise regression. The paper considers over-fitting and multi-collinearity to adjust LARS. SRF use forward selection process and greedy rule which stand on the basis of correlation between input vector and residual. We can see the disadvantage of SRF in counter example of existing research. SRF exclude non-selected input variable correlated with selected input variable. So it can't consider various subset of the input variable [3].

Improved algorithms for this disadvantage of Stepwise Regression are Stagewise, Lasso [6], LARS (Least Angle Regression, the "S" suggesting “Lasso” and “Stagewise”) and so on. Stagewise and Lasso can be transformed into LARS, this transformation proved mathematically. That means these algorithms have same prediction accuracy. And calculative performance of LARS is better than that of Stagewise and Lasso [5]. Therefore, we focus on the LARS. LARS uses correlation between correlation between input vector and residual. This measure equals to that of Stepwise regression. But LARS improves prediction accuracy, using different direction select and step length. So LARS effects on counter example of SRF [3].

However, LARS has some fatal disadvantage. [5] shows disadvantages. Discussion of Robert A. Stine shows that if there is multi-collinearity between input variables, LARS is exposed to serious over-fitting problem. In real situation, this problem effects on fault variable selection. In generally, we use as many candidate input variables as possible to improve accuracy of model. Increasing number of candidate input variables bring about high chance to multi-collinearity. In this case LARS has high chance to making unstable prediction model.

We judge that one of the causes which make the problem described in the preceding paragraph is under-estimation of correlated input variables. More precisely, as iterations progress and some input variables are selected, part of the explanation made of selected variables is removed from residual vector which is measure of selecting candidate variables. Then the selected variables and care for other variables that have a high rating is relatively lower. And we improve the problem in that point of view.
4. MAIN IDEA

The problem which is improved in the paper exist both 'SRF' and LARS. And explain improvement of SRF is easier to understand visually. So I will explain the problem of the SRF and improving method.

We assume that three candidate input variables is considered to create prediction model for one output variable y. Process for selecting variables used in SRF is described in Figure 1.

![Figure 1: Process of SRF](image)

In figure 1, each input variable vectors and output variable vector are expressed with ellipses of same size. Overlap of two ellipses is the vector projected on another vector. In addition, divide the overlap area by ellipse area is absolute value of correlation between two vectors. In this point, proposed algorithm is as described below. Figure of step 1 describe comparison between each of correlations between output variable and input variables. In Figure 1 overlap area between \( X_1 \) and \( y \) is largest. So \( X_1 \) is selected and constitute \( S \). In step 2 and step 3, characteristic of algorithm is well brought out. In step 2 \( y' \) is projection vector \( y \) on \( X_2 \). Next we calculate absolute value of correlations between \( y' \) and input variable left (\( X_1, X_3 \)). In this step, the variable which has much overlap area with \( X_2 \), is underestimated. As a result, we can't select correct variable and accuracy of prediction model will be low.

These problems can be solved in the same way as the Figure 2. When we calculate correlations, use translated vectors \( X_1' \) and \( X_3' \) instead of original vector \( X_1 \) and \( X_3 \). \( X_1' \) is projection vector \( X_1 \) on \( X_2 \). \( X_3' \) is made in same way. In this way, we can calculate more accurate measure using correlation.

![Figure 2: Process of proposed algorithm](image)

We modify a pseudocode of SRF as described below.

**Pseudocode of Stepwise Regression Forward**

- **Initialization**
  A. Initialize residual vector as output vector \( Y \)
  B. Initialize \( S, N \) (set of input vector) as described below
    \[
    S = \{ \}, \quad N = \{ i \mid i \text{ is index of input variable } X_i \} \]
  C. Set \( F \) to \( \text{Enter} \) value which is a selection criterion
  D. Calculate initial residual value as below. \( n \) is number of data, \( p \) is iteration number.

\[
R_p = \sum_{i=1}^{n} Y_i^2
\]  

(3)
• **Step 1**
  A. Calculate correlations between $X_i$ and residual vector. $X_i$ is an element of $N$.

  \[
  Corr(X_i, \text{Residual}) = \frac{\sigma_{X_i, \text{Residual}}}{\sigma_{X_i}, \sigma_{\text{Residual}}}
  \]  
  (4)

  B. Find $X_{\text{max}}$ which have maximum absolute correlation.

• **Step 2**
  A. Set residual vector as (5).

  \[
  \text{Residual} = \text{Residual}_{\text{old}} - \frac{\text{Residual}_{\text{old}}^T X_{\text{max}}}{X_{\text{max}}^T X_{\text{max}}} X_{\text{max}}
  \]  
  (5)

  B. Calculate residual value as (6).

  \[
  R_{p+1} = \sum_{i=1}^{n} \text{Residual}_i^2
  \]  
  (6)

• **Step 3**
  Calculate $F - \text{Statistics}$ as (7). Where $k$ = number of selected variable, $|S|$

  \[
  F - \text{Statistics} = \frac{R_p - R_{p+1}}{R_{p+1}} (n - k - 2).
  \]  
  (7)

  If $F - \text{Statistics}$ is lower than $F - \text{to} - \text{Enter}$
  A. Set $S$, $N$ as (8).

  \[
  S^+ = X_{\text{max}}, \quad N^- = X_{\text{max}}
  \]  
  (8)

  B. $R_p = R_{p+1} + 1$

  C. Go to Step 1

  If $F - \text{Statistics}$ is higher than $F - \text{to} - \text{Enter}$, complete algorithm.

A pseudocode of the new algorithm, the modifications to the code as described above. In step 1, change (a) as below.

  A. Calculate correlations between $X_{i'}$ and residual vector. $X_{i'}$ is an element of $N$. In first iteration, $X_{i'}$ equal to $X_i$.

  \[
  Corr(X_{i'}, \text{Residual}) = \frac{\sigma_{X_{i'}, \text{Residual}}}{\sigma_{X_{i'}}, \sigma_{\text{Residual}}}
  \]  
  (9)

And in step 3, add below process after B.

  C. $X_{i'} = X_i - \frac{X_i^T X_{\text{max}}}{X_{\text{max}}^T X_{\text{max}}} X_{\text{max}}, \quad i \in N$

  (10)

Calculation direction method of LARS same as that of SRF. In other words, calculate correlations between residual vector and input vectors. In this time, LARS only make orthogonal selected input vector on residual vector but do not on non-selected input vector. And select input variable whose correlation is largest. So the same problem described above, occur in LARS. The non-selected vector which highly related with output variable and selected input vectors will be underestimated in calculating correlation. As a result, it may not be the correct choice of variables.

To adjust this problem add the formula (10) to pseudocode LARS, and after then use $X'$ instead of $X$. This modified algorithm will call Adjusted LARS. Pseudocode of Adjusted LARS is described follow.
Pseudocode of the Adjusted LARS

- **Initialization**
  A. Normalize input vector $X_i$ (for $i = 1, \cdots, p$), output vector $Y$ by mean and variance
  B. Initialize $S_N$ (set of input vector) as described below
    
    $$ \mathcal{S} = \{\}, \quad \mathcal{N} = \{i \mid i \text{ is index of input variable } X_i\} $$
  C. Initialize residual vector as output vector $Y$
  D. Initialize sign variable $S_i$ (for $i = 1, \cdots, p$) as 0. $S_i \in \{-1, 0, 1\}$

- **Step 1**
  A. Calculate correlations between $X_i'$ and residual vector (9). $X_i$ is an element of $N$. In first iteration, $X_i'$ equal to $X_i$.
  B. Select input variable $\text{Corr}_{\max}$ which has maximum correlation.
    
    $$ \text{Corr}_{\max} = \max_{i=1}^n (\text{Corr}_i) $$
  C. Save sign of correlation between $X_i'$ and residual vector in $S_i$.
  D. Set $S_N$ as (8).
  E. Define vector $\mathbf{X_s}$ as (12)
    
    $$ \mathbf{X_s} = (\cdots, s_0, x_0, \cdots)_{i \in S}, s_i = \pm 1 $$

- **Step 2**
  A. Set residual vector as (13).
    
    $$ \text{Residual} = \text{Residual}_{old} - \frac{\text{Residual}_{old}^{T} \mathbf{X'}_\max \mathbf{X'}_\max}{\mathbf{X'}_\max^{T} \mathbf{X'}_\max} \mathbf{X'}_\max $$
  B. Calculate residual value as (6).
  C. Calculate as follow.
    
    $$ A_S = [I_S^{T} (X'_{S}^{T} X'_{S})^{-1} 1_S]^{-\frac{1}{2}} $$
    
    $$ W_S = A_S (X_{S}^{T} X'_{S})^{-1} 1_S $$
    
    $$ u_S = X_S W_S $$
    
    $$ a_N = X_N^{T} u_S $$
    
    1s is vector whose all elements are 1 and size is $p$. And where $j \in N$
    
    $$ \gamma^+_j = \frac{\text{Corr}_{\max} - \text{Corr}_j}{A_S - a_{N_j}}, \gamma^-_j = \frac{\text{Corr}_{\max} + \text{Corr}_j}{A_S + a_{N_j}} $$
    
    $$ \gamma_j = \max(\gamma^+_j, \gamma^-_j), \ \gamma = \min(\gamma_j) $$
    
    $$ X_{\max} = X_j, \ \min(\gamma_j) = \gamma_j $$
    
    If $\gamma > 0$, $s_i = +1$ else $s_i = -1$
    
    $$ \text{Residual} = \text{Residual} - \gamma A_j, \ \text{Corr}_j = \text{Corr}_j - \gamma A_j, \ j \in N $$

- **Step 3**
  Repeat Step1, Step2 until satisfy stopping criterion (defined by user)

The Adjusted LARS different from LARS in searching direction. In figure3 comparing Adjusted LARS with LARS, the Adjusted LARS use $X_{\cdot}'$ (from Formula 10) instead of $X_{\cdot}$. If we use it, the point of same correlation is pulled if $X_{\cdot}$ and $X_{\cdot}'$ is not orthogonal. Because underestimation of $X_{\cdot}'$ which is generated by correlation between $X_{\cdot}$ and $X_{\cdot}'$, is reduced.
5. EXPERIMENT

To compare new algorithm with LARS, we make an experiment with simulation data. The data is generated by the rules described in [7]. It generates two classes of data. We set the correlation (between input variables) of the one class to 0, and correlation of the other to 0.9.

We generate data as follow. The number of rows (the number of candidate input variables) \( m \) is 400. And the number of column (the number of data) \( n \) is 800. Input variable \( x_i \) is generated by multivariate normal distribution and satisfies i.i.d. \( N(0,1) \). Covariates of \( x_i \), for \( i = 1, \cdots, n \), were generated independently from a multivariate normal distribution with zero mean and with covariance satisfying \( E(x_i, x_j) = \rho^{ij} \). Experiment independently \( \rho \) is 0 or 0.9. And white noise is normally distributed whose mean is 0, standard deviation is 1. \( \beta_j \) coefficients were in 15 clusters of 7 adjacent variables centered at every 25th variable. The value is calculated (14).

\[
\beta_{25a+j} = | h - j |^{25} | j < h, h = 4_i (\alpha = 1, 2, \cdots, 15) 
\]

As a result, we generate 105 variables whose \( \beta \) is not 0, and generate 295 variables whose \( \beta \) is 0. After then, to adjust theoretical \( R^2 \) value to 0.75, \( \beta_j \) took additional compensation by multiplying the specific value [7]. Although the number of variables whose \( \beta \) is not 0 is changed, this type of simulation has the advantage by considering different configurations for the dimension \( m \), sample size \( n \), correlation \( \rho \) and the number of nonzero coefficients [8]. For each \( \rho \) correlation setting simulations were repeated 100 times independently. Results are recorded in Table1.

<table>
<thead>
<tr>
<th>( \rho = 0 ) (uncorrelated X)</th>
<th>( \rho = 0.9 ) (correlated X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho \text{ pe}(\mu) )</td>
<td>( \rho \text{ pe}(\mu) )</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>SRF 135.53 0.876 70.35 0.367 0.075 129.24 0.884 137.10 0.552 0.208</td>
<td></td>
</tr>
<tr>
<td>LARS 210.69 0.907 126.63 0.547 0.055 99.51 0.962 75.77 0.347 0.135</td>
<td></td>
</tr>
<tr>
<td>Adj-LARS 215.70 0.911 60.77 0.321 0.000 130.31 0.966 62.34 0.324 0.000</td>
<td></td>
</tr>
</tbody>
</table>

We must select variables whose \( \beta_j \) is not 0 (true variable) and must not select variables whose \( \beta_j \) is 0 (false variable). TotalMiss is the total number of misclassified variables. FDR is the false discovery rate which is the rate false variable is selected. FNR is the false nondiscovery rates defined as the rate true value is not selected. To represent
TotalMiss, FDR, FNR values we use the averaged values from the 100 simulations. The 'M' recorded in the table is the average number of variables selected by a procedure. The performance value (prediction power) \( pe(\hat{y}) \) (15), again averaged over the 100 simulations [8].

\[
pe(\hat{y}) = 1 - \frac{\| \hat{y} - y \|^2}{\| y \|^2}
\]

(15)

In this experiment each three algorithm use \( C_p \) criterion. \( C_p \) can be calculated as (16).

\[
C_p(\hat{y}_k) = \frac{\| y - \hat{y}_k \|^2}{\sigma^2} - n + 2k
\]

(16)

The overall conclusion from [Table 1] is follow.

- In Adjusted LARS, Total-Miss, FDR, FNR is lower than SRF and LARS. It means that fault variable selection is reduced. Especially in both case, FNR is 0. In this experiment, the Adjusted LARS doesn't miss meaning variable.

- In correlated case, \( m \) of LARS is small than Adjusted LARS. And FDR of LARS is similar to Adjusted LARS'. But FNR of LARS is higher than Adjusted LARS'. And in uncorrelated case FDR is high. It means that LARS throws correlated input variables rapidly and Adjusted LARS makes decisions more carefully.

- A performance of the Adjusted LARS is better than other two algorithms.

6. Conclusion

We presented the Adjusted LARS algorithm to improve accuracy of prediction model. We consider multi-collinearity between input variables, and to improve algorithm we remove correlated factor between input variables by creating orthogonal vector. To make orthogonal vector we use QR factorization in algorithm [9].

In the experiment, we can see the results were better than the original LARS algorithm. The number of misclassified variables is decreased. Especially, the number of throwing meaning variable is reduced a lot. But we must to prove that the results of experiment with different open data and real process data. And verification selection criteria for further research should be. For example AIC (Akaike information criterion), BIC (Bayesian Information Criterion), \( C_p \) can be used. And to improve algorithm more correct, we must consider over-fitting problem. One of the method to avoid over-fitting is n-fold cross-validation. To apply n-fold cross-validation we must to define rule to select variables which are selected in each fold.

Reference

A Tabu Search approach for pallet transport management problem

Pierpaolo Caricato*, Antonio Grieco, and Francesco Nucci

Department of Innovation Engineering
University of Salento
Lecce, 73100, Italy

ABSTRACT

In Flexible Manufacturing System, parts move among the resources in order to be processed. Usually, a classical pallet transport system adopts the railway carrier approach. Consequently, the transport layout is linear and bidirectional. Each time a pallet/part has to be moved in the system, a transport trip is issued and waits to be satisfied by the transport system. The pallet transport management decides the sequence the transport trips have to be executed. In general, the objective consists in minimizing the makespan to perform all the trips. When high rate values of the parameter “number of transport trips per time unit” are required, the transport system can become the system bottleneck. This leads to starvation and blocking on machining resources due to parts waiting to be moved by transport system. In such a case, the problem is multi-objective. The main objective consists in the minimization of the idle resource time due to transport system; whereas the minimization of the transport makespan is secondary. The solution consists in determining the exact list of pallet trips performed by the transport system. The considered approach proposes a Tabu Search based approach to solve the illustrated problem. In particular, in order to solve efficiently the problem, we considered the list of trips to be performed by the vehicle as a solution. The validation of the method is performed on a case study in order to compare the performance with state-of-the-art approaches.

1. INTRODUCTION

Recently, new ideas related with the design of manufacturing system architectures provided with the correct level of flexibility in order to face the specific production problem has been introduced. Such approaches tune system flexibility on the production problem, in order to cope with the uncertainty related to product evolution [1]. This problem leads to develop a methodology to analyze the requirements in terms of flexibility of the firm, and to design the system starting from the flexibility levels required. Consequently, new manufacturing systems, with the minimum level of flexibility required, have to be designed. Considering a classical Flexible Machining System (FMS), if the system flexibility is reduced, some CNC machines can be replace by dedicated machines not able to perform the entire part program. Consequently, parts need processing operations on different machines. As a result, a relevant management question consists in configuring the transport system in order to face the higher rate of pallet transfers. In FMS, a classical transport solution consists in a vehicle carrier. If the pallet transfer rate increases, the management policy has to avoid the transport system becomes the system bottleneck. Indeed, if pallet trip is not performed in time, a machine gets idle. Such a problem is common in other fields (sewing, assembling) in which resources are placed in a linear layout and a vehicle is adopted as transport system.

Managing a transport system efficiently is an important task [2]. In [3] a review on design and control of automated guided vehicle systems is presented. Various key related issues are addressed including guide-path design, determining vehicle requirements, vehicle scheduling, idle-vehicle positioning, battery management, vehicle routing and deadlock resolution. In [4], Kim and Tanchoco propose an algorithm based on Dijkstra’s shortest-path method to schedule vehicles based on the nodes’ time windows. This approach produces a deadlock-free schedule. However, as noted by the authors, a small change in the schedule may destroy it completely. Taghaboni and Tanchoco, see [5], introduce an incremental route planning and scheduling algorithm. Other approaches are introduced in [6] and [7]. More details about the vehicle routing issue can be found in [8].

* Corresponding author: Tel.: (0039) 0832-297806; Fax: (0039) 02-700441811; E-mail: pierpaolo.caricato@unile.it
On one hand, the Vehicle Routing Problem (VRP) has been widely investigated, in order to design least cost routes from one depot to a set of geographically scattered points. On the other, the scheduling of vehicle tasks received less attention and has always been classified as a special case of VRP [9].

The problem considered in this work deals with the management of the transport system associated with a job shop production plant. Such a plant requires pallets to be moved by a vehicle along a railway system passing through inline disposed machines. Parts are processed by machines in accordance to a given part program and are transferred from a machine to the following one. Each part request to be transferred from a station to another along the line is referred as Part Trip (PT). Transport system manager has to schedule the PT in order to allow the best use of the transport vehicles in the system. This lead to: minimizing empty travel performed by the vehicle between two consecutive PTs. The optimal use of the transport system allows reducing machine idle time (due to lack of parts).

The remainder of the paper is organized as follows. In the next section, we describe the problem statement and its mathematical formulation. Then, we formulate the proposed approach adopting Tabu Search technique. In the last section, we report the experimental campaign and the corresponding results we obtained to validate our approach.

2. PROBLEM STATEMENT

The problem treated considers a set of tasks (or trips) characterized by an origin, a destination and a duration. The tasks are given a priori and are composed of one or several activities. Between two successive tasks there is an inter-task transit time period of a certain duration and cost. The problem to be solved consists of forming a schedule, i.e., sequences of tasks where each task is performed exactly once. This type of problem is encountered in several fields, such as airline, rail, schoolbus and urban transportation. Dantzig and Fulkerson [10] were the first to formulate a fixed schedule problem as a minimum cost flow circulation problem.

An example of system layout is reported in Figure 1. The resources are placed in line along the path of the vehicle according to an assigned distance. The speed vehicle is assumed to be constant.

![Figure 1: System layout](image)

The Vehicle Scheduling Problem with Time Windows and a unique vehicle (VSPTW) can be formulated as Traveling Salesman Problem with Time Window (TSPTW) considering the trips to be performed as nodes of a graph. Consider a set of \( n \) trips \( \{ T_1, T_2, ..., T_n \} \), where trip \( T_i \) has a given duration \( \lambda_i \) and \( \delta_i \) is the due date in which \( T_i \) has to be completed. Consider also an initial position in which the vehicle is stationed. Let the set of nodes \( N = \{1, 2, ..., n\} \) represent the set of trips, and node \( \{n + 1\} \) represent the single depot. Let also \( t_{ij} \) denote the travel time between the ending point of trip \( T_i \) and the starting point of trip \( T_j \). Note that in general \( t_{ij} \neq t_{ji} \). For simplicity, we will assume that \( t_{ij} \) includes the duration \( \lambda_i \) of the trip \( T_i \).

Let \( c_{ij}, i,j \in N \) be the cost incurred if a vehicle performs trip \( T_j \) immediately after trip \( T_i \). This cost is usually a function of the distance between the ending point of trip \( T_i \) and the starting point of trip \( T_j \). It can also include the fixed cost incurred to cover one of the trips, say trip \( T_i \). Let next \( c_{n+1,j} \) be the cost incurred if the vehicle undertakes trip \( T_j \) as the first trip after leaving the depot. Similarly, the cost \( c_{j,n+1} \) is incurred when trip \( T_j \) is the last trip before the vehicle returns to the depot. This cost can also include the fixed cost incurred to cover trip \( T_j \). The problem consists of finding an assignment of vehicles to trips in such a way that:
A Tabu search approach for pallet transport management problem

- each trip is covered exactly once by a vehicle;
- vehicle leaves from the depot, covers a sequence of pairwise compatible trips and returns to the depot at the end of the route;
- the sum of the costs of the routes traveled by the vehicles used in the solution is minimized;
- each trip has to be completed before a given due date.

By making duplicates of the depot node, \( o \) and \( d \) for the origin and the destination of any feasible route, it is simple to describe the path origin and destination. This allows checking if the due date constraints are enforced. To formally describe the problem, define the graph \( G = (V, A) \), where \( V = N \cup \{o,d\} \) is the set of nodes and \( A \) is the set of arcs as described in (1). Consequently, the number of arcs is equal to (2).

\[
A = \{(o,j) | j \in N\} \cup \{(i,j) | i,j \in N, i \neq j\} \cup \{(i,d) | i \in N\}
\]

(1)

\[
|A| = n + n(n - 1) + n = 2n + n^2 - n = n^2 + n = n(n + 1)
\]

(2)

Let the variables \( x_{ij} \) be the binary flow on arc \((i,j) \in A\): \( x_{ij} = 1 \), if a vehicle covers trips \( T_j \) after trip \( T_i \), and \( x_{ij} = 0 \), otherwise. Moreover, the variables \( \tau_i \) specifies the start of service at node \( i \), \( i \in N \cup \{o,d\} \).

\[2.1. \text{Formulation} \]

The model can be formulated as a pure network flow problem:

Minimize

\[
\sum_{(i,j) \in A} c_{ij}x_{ij}
\]

subject to:

\[
\sum_{j \in N \cup \{d\}} x_{ij} = 1 \quad \forall i \in N
\]

(4)

\[
\sum_{j \in N} x_{oj} = 1
\]

(5)

\[
\sum_{i \in N \cup \{o\}} x_{ij} - \sum_{i \in N \cup \{d\}} x_{ji} = 0 \quad \forall j \in N
\]

(6)

\[
\sum_{i \in N} x_{id} = 1
\]

(7)

\[
x_{ij}(\tau_i - t_i - \tau_j) \leq 0 \quad \forall (i,j) \in A
\]

(8)

\[
\tau_i + \lambda_i \leq \delta_i \quad \forall i \in N
\]

(9)

\[
x_{ij} \in \{0,1\} \quad \forall (i,j) \in A
\]

(10)

\[
\tau_i \geq 0 \quad \forall i \in N \cup \{o,d\}
\]

(11)
It is a nonlinear program with $O((n + 2)^2)$ variables and $O((n + 2)^2)$ constraints.

The objective function (3) seeks to minimize the sum of travel. Constraints (4) ensure that each trip is performed exactly once, while constraints (5)-(7) are flow conservation constraints. Constraints (8)-(9) ensure feasibility of the time schedule. Note that the sub-problem defined by (4)-(7) and (10) is a minimum cost flow problem (in fact an assignment problem with $(n + 2)$ rows and $(n + 2)$ columns). Whereas, the additional constraints (8), (9) and (11) deal with the time schedule. It is important to note that sub-tours are eliminated by constraints (8). Using a big $M$ constant, such constraints (8) can be linearized and rewritten as (12).

\[ t_i + t_{ij} - t_j \leq M(1-x_{ij}) \quad \forall (i,j) \in A \]  

(12)

Note that the large constant $M$ can be replaced by $M_{ij} = \delta_i + t_{ij}, (i, j) \in A$. In this way, it is possible to use (13) instead of (8).

\[ t_i + t_{ij} - t_j \leq (\delta_i + t_{ij})(1-x_{ij}) \quad \forall (i,j) \in A \]  

(13)

Research on the TSPTW has been scant. In [9] and [11], it is proved that such a problem is NP-hard. Anyway, considering the linear layout case, researches have investigated the vehicle routing problems on a line-shaped network [12]. The solution methods for the VRPTW have been studied in [13] and [14]. In particular the most interesting approaches that are suitable to be applied in our case are in the following.

Homberger and Gehring [15] present two evolution strategies, together with GAs and evolutionary programming, the evolution strategies form the class of evolutionary algorithms. By definition, the main differences between these three types of algorithms lie in the representation and in the role of mutation. In Evolution Strategies (ES) of [15] the individual representation includes a vector of so-called “strategy parameters” in addition to the solution vector and both components are evolved by means of recombination and mutation operators these strategy parameters refer to how often a randomly selected local search operator is applied. Bent and Van Hentenryck [16] present a Two Stage Hybrid Metaheuristic (TSHM), where in the first stage a basic simulated annealing is used to minimize the number of routes, and the second stage focuses on distance minimization using the Large Neighborhood Search. Gambardella et al. [17] use an Ant Colony Optimization (ACO) approach that minimizes the total traveled distance.

3. Tabu Search Approach

Tabu search (TS) is a local search procedure that iteratively moves from a current solution to its best neighbor even if this causes deterioration in the objective function value [18] in order to avoid being locked into local minimums. To avoid cycling, a short term memory (like a tabu list or tabu attributes) is used. In addition, a diversification can be used to visit new regions of the solution space, whereas intensification is used to explore the neighborhood of the current solution.

The permutation of the trips to be accomplished has to be determined in order to optimize the system’s performances. Given a certain permutation, this can be adopted as the current solution for the TS algorithm. In particular, a generic solution for the TS algorithm is represented by a permutation of the indexes in the afore-defined set $N$.

3.1. Initial Solution

TS needs a feasible initial solution in order to start exploring the solution space. This first solution is determined through a greedy constructive algorithm. At each stage, the algorithm selects the unselected trip nearest to the current trip. We choose among the trips to be performed, the one having the starting point that is nearer to the initial position of the vehicle. The next trip to be performed is selected, among the remaining trips, considering the minimal cost required to move from the arrival station of the currently selected trip to the departure station of the candidate trip. Using the notation introduced before, this can be formalized as:

List $M = \emptyset$
prec $= o$

for (j=1 to n)
A Tabu search approach for pallet transport management problem

\[ m_j = \arg \min_{i \in N \cdot M} \{c_{\text{prec},i}\} \]
\[ M.\text{add}(m_j) \]
\[ \text{prec} = m_j \]

where \( M = (m_1, m_2, ..., m_n) \) is the initial solution for the TS: \( m_1 \) is the first trip to be performed, \( m_2 \) the second one and so on.

3.2. MOVE AND NEIGHBORHOOD

Once an initial solution has been determined, the following solutions to be explored are determined through moves, i.e. defining a way to determine a new feasible solution from the current one. In the proposed approach a generic move is determined using a 2-opt approach [19]. Hence, the neighborhood of the current solution is given by the set of all possible swaps that can be made. The size of the neighborhood is hence given by (14).

\[ \binom{n}{2} = \frac{n \cdot (n - 1)}{2} \]  \hspace{1cm} (14)

Each solution in the neighborhood has to be tested for feasibility. With the proposed move, indeed, it is not possible to guarantee the feasibility of the new solution, since the respect of the due date constraint is not necessarily satisfied. A measure \( \Phi \) of the infeasibility of a given solution can be defined, using (9), as reported in (15).

\[ \Phi = \max_i \{0, \tau_i + \lambda_i - \delta_i\} \]  \hspace{1cm} (15)

The evaluation of the moves in the neighborhood is determined calculating the delta-cost between the current solution and the new solution defined by the selected swap. Let \( i \) in \( N \) be a generic trip: we define the before(\( i \)) and after(\( i \)) functions that indicate in the current solution, respectively, the trips to be performed before and after \( i \). Let \( i \) and \( j \) in \( N \) be the trips to be swapped, the delta-cost is equal to (16).

\[ \Delta = [c_{\text{before}(i), j} + c_{\text{after}(j), i} + c_{\text{before}(j), i} + c_{\text{after}(j), i}] - [c_{\text{before}(i), i} + c_{\text{after}(i), i} + c_{\text{before}(j), j} + c_{\text{after}(j), j}] \]  \hspace{1cm} (16)

where the former term is the contribution of the trips involved in the swap to the overall cost function in the new solution, while the latter is the same contribution in the current. The 2-opt approach is described in Figure 2.

\[ \begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{move_2opt.png}
\caption{Move with 2-opt approach and delta-cost calculus}
\end{figure} \]

Once all moves have been performed, only one has to be chosen. If at least one of the candidate new solutions are feasible, then the selected solution is the one having the minimum \( \Delta \). On the contrary, if all candidate solutions are infeasible, then the one having the minimum \( \Phi \) is selected.

3.3. TABU LIST AND DIVERSIFICATION

In order to avoid local minimums, TS uses a tabu list that temporarily forbids certain moves. In the considered approach, each move is characterized by two important attributes: namely, the two trips involved in the selected swap. Both indices of the applied swap are stored in the tabu list; hence none of them can be selected for a swap for a given number of iterations.

Diversification is used in order to escape from local minimums in the cases in which the tabu list approach is not sufficient. A maximum number of iterations without improvement in the objective function is determined. When this occurs, the solution is changed using a diversification approach. In the proposed algorithm, we diversify using a 3-
opt approach [12]. We consider every possible sequence of three adjacent trips (hence \( n-2 \) triplets) in the current solution and we generate the two candidate solutions that are given by the permutations of the three selected trips. We then select the best (or the least unfeasible) solution among the generated ones, see Figure 3.

![Figure 3: Diversification with 3-opt approach.](image)

4. CASE STUDY

4.1. EXPERIMENTAL CAMPAIGN

The mechanism for the generation of case studies has been developed and tuned on the basis of an experimental campaign performed on industrial case studies, having the linear layout transport system. In particular, we observed cases in which the necessity the high transport rate leads to resource idle times. The case studies have been generated considering three different parameters:

1. Number of trips to be scheduled (\( n \))
2. Number of stations in the linear path
3. Due date looseness

The number of task to be scheduled directly influences the problem complexity. Since, the higher is the number of stations, the higher is the number of trip requests, also the second parameter affects the complexity. The third parameter describes the cardinality of the feasible solutions. If due date looseness is high, several feasible solutions exists and it is possible to find the optimal one minimizing the overall cost. Otherwise, critical trips force the solution algorithm to schedule them early to avoid lateness.

Due dates have to be carefully generated in order to avoid empty domain sets. The generation algorithm for trip due dates we adopted is the following:

1. Define a looseness parameter \( \alpha \) between 0 and 1
2. Calculate \( D_1 \) as the average of trip durations
3. Calculate \( D_2 \) as the average of travel time between two trips
4. Calculate the estimate of the total trip duration as \( (D_1 + D_2) \cdot n \)
5. Set the due date \( d_i = (D_1 + D_2) \cdot (1 + \alpha) \cdot \rho_i \) where \( \rho_i \) is a random number between 0 and 1 (uniform distribution).

This method calculates an estimation of the total duration trip considering the average values of the transition times and trip length times. The estimation is increased by a looseness factor \( \alpha \) (e.g. 20%). Due dates are randomly generated between 0 and the estimation. This method well approximates the distribution of real due date data in which low due date values lead to high priority trips.

The considered test case have been generated by considering the combination between the number of trips \( n \) in the set \{20, 40, 60, 80, 100\} and the due date looseness \( \alpha \) in the set \{0, 0.25, 0.5, 0.75, 1\}. The number of stations is
set to n/2+2. The initial vehicle position is set to the middle station of the linear path. For each experiment, a time limit of n/10 minutes has been established. The proposed Tabu Search (TS) approach has been compared with the three methods presented at the end of section 2: Evolution Strategies (ES) [15], Two Stage Hybrid Metaheuristic (TSHM) [16] and Ant Colony Optimization (ACO) [17].

### 4.2. Computational Results

The experimental campaign has been performed on a PC Intel Core 2 CPU 2.33 GHz 1GB RAM with Microsoft Windows XP Pro SP3. Results are reported in Table 1. In columns 4-7, the objective value of the best solution found, using one of the methods, is reported. In the last three columns the gap (%) between the TS solution and one of the three benchmarks is reported. If the gap is positive, the TS solution is better than the other one.

For each experiment, at least a feasible solution exists. In the first five experiments, all the methods are able to find the same optimal solution. In such cases (exp. 1-5) we were also able to solve the MIP problems obtained using expressions (12) and (13) instead of (8) by a commercial MIP solver (ILOG CPLEX 11). However, in the remaining experiments 6-25, the MIP solver could not even find an initial feasible solution within the fixed time limit. As the number of trips increases and the due dates looseness decreases, the TS approach is able to find a better result than the other ones. Indeed, on one hand the average gaps are 3%, 1.2% and 3%. On the other, if only experiments 21-25 are considered, the average gap values are 3.5%, 2.5% and 4.0%. The good performance of TSHM shows the capabilities of hybrid metaheuristics to be adapted to our problem.

<table>
<thead>
<tr>
<th>EXP</th>
<th>n</th>
<th>α (%)</th>
<th>TS</th>
<th>ES</th>
<th>TSHM</th>
<th>ACO</th>
<th>GAP TS-ES</th>
<th>GAP TS-TSHM</th>
<th>GAP TS-ACO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>0</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>25</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>50</td>
<td>142</td>
<td>142</td>
<td>142</td>
<td>142</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>75</td>
<td>136</td>
<td>136</td>
<td>136</td>
<td>136</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>100</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>0</td>
<td>448</td>
<td>486</td>
<td>454</td>
<td>470</td>
<td>8.5%</td>
<td>1.3%</td>
<td>4.9%</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>25</td>
<td>436</td>
<td>450</td>
<td>440</td>
<td>465</td>
<td>3.1%</td>
<td>1.0%</td>
<td>6.7%</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>50</td>
<td>432</td>
<td>452</td>
<td>424</td>
<td>460</td>
<td>4.7%</td>
<td>-1.8%</td>
<td>6.5%</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>75</td>
<td>425</td>
<td>454</td>
<td>424</td>
<td>450</td>
<td>6.8%</td>
<td>-0.2%</td>
<td>5.9%</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>100</td>
<td>412</td>
<td>417</td>
<td>408</td>
<td>441</td>
<td>1.2%</td>
<td>-1.0%</td>
<td>7.1%</td>
</tr>
<tr>
<td>11</td>
<td>60</td>
<td>0</td>
<td>751</td>
<td>770</td>
<td>767</td>
<td>759</td>
<td>2.6%</td>
<td>2.1%</td>
<td>1.0%</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
<td>25</td>
<td>738</td>
<td>758</td>
<td>758</td>
<td>759</td>
<td>2.7%</td>
<td>2.6%</td>
<td>2.8%</td>
</tr>
<tr>
<td>13</td>
<td>60</td>
<td>50</td>
<td>729</td>
<td>746</td>
<td>745</td>
<td>757</td>
<td>2.3%</td>
<td>2.2%</td>
<td>3.8%</td>
</tr>
<tr>
<td>14</td>
<td>60</td>
<td>75</td>
<td>715</td>
<td>775</td>
<td>732</td>
<td>752</td>
<td>8.4%</td>
<td>2.4%</td>
<td>5.2%</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
<td>100</td>
<td>702</td>
<td>721</td>
<td>698</td>
<td>714</td>
<td>2.7%</td>
<td>-0.6%</td>
<td>1.8%</td>
</tr>
<tr>
<td>16</td>
<td>80</td>
<td>0</td>
<td>1254</td>
<td>1297</td>
<td>1300</td>
<td>1267</td>
<td>3.5%</td>
<td>3.7%</td>
<td>1.0%</td>
</tr>
<tr>
<td>17</td>
<td>80</td>
<td>25</td>
<td>1247</td>
<td>1297</td>
<td>1279</td>
<td>1259</td>
<td>4.0%</td>
<td>2.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>18</td>
<td>80</td>
<td>50</td>
<td>1235</td>
<td>1266</td>
<td>1239</td>
<td>1255</td>
<td>2.5%</td>
<td>0.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>19</td>
<td>80</td>
<td>75</td>
<td>1221</td>
<td>1237</td>
<td>1259</td>
<td>1248</td>
<td>1.3%</td>
<td>3.2%</td>
<td>2.2%</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>100</td>
<td>1192</td>
<td>1227</td>
<td>1204</td>
<td>1229</td>
<td>2.9%</td>
<td>1.0%</td>
<td>3.1%</td>
</tr>
<tr>
<td>21</td>
<td>100</td>
<td>0</td>
<td>1798</td>
<td>1848</td>
<td>1869</td>
<td>1869</td>
<td>2.8%</td>
<td>3.9%</td>
<td>3.9%</td>
</tr>
<tr>
<td>22</td>
<td>100</td>
<td>25</td>
<td>1776</td>
<td>1853</td>
<td>1816</td>
<td>1847</td>
<td>4.4%</td>
<td>2.3%</td>
<td>4.0%</td>
</tr>
<tr>
<td>23</td>
<td>100</td>
<td>50</td>
<td>1754</td>
<td>1834</td>
<td>1789</td>
<td>1810</td>
<td>4.6%</td>
<td>2.0%</td>
<td>3.2%</td>
</tr>
<tr>
<td>24</td>
<td>100</td>
<td>75</td>
<td>1732</td>
<td>1793</td>
<td>1783</td>
<td>1801</td>
<td>3.5%</td>
<td>2.9%</td>
<td>4.0%</td>
</tr>
<tr>
<td>25</td>
<td>100</td>
<td>100</td>
<td>1697</td>
<td>1737</td>
<td>1718</td>
<td>1783</td>
<td>2.4%</td>
<td>1.2%</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

| average | 3.0% | 1.2% | 3.0% |
5. SUMMARY

In this work, we proposed a Tabu Search approach to solve the vehicle scheduling problem in linear layout system. We developed our approach by defining the initial solution, the move, the neighborhood, the tabu list and diversification. In particular, the proposed approach implementation reflects the characterization of the scheduling problem in the linear layout. The benchmark of the considered approach has proven its effectiveness against state-of-the-art approaches.

REFERENCES

Lean Automated Manufacturing – A Panacea or a Potential Path to Ruin?

Hongyi Chen*, Richard R. Lindeke1*, and David A. Wyrick2

1 Department of Mechanical & Industrial Engineering
University of Minnesota Duluth
Duluth, MN 55812
USA

2 Department of Industrial Engineering
Texas Tech University
Lubbock, TX 79409-3061
USA

ABSTRACT

In times of severe economic challenge, the natural reaction is to make decisions that can make an organization become as lean and focused as possible. This reaction of implementing “lean to the max” may, in fact, provide outcomes that are exactly the opposite of what is desired. This paper discusses the benefits and pitfalls associated with lean manufacturing management. These benefits and pitfalls may be seen as outcomes based on the degree to which lean is implemented. Elements of lean discussed in this paper address organizational waste, human resources, distributed design, supply chain management, customer management, and the financial system. Potential solutions and recommendations are made to help organizations remain focused on being centered as well as lean.

1. INTRODUCTION

The benefits of Lean Enterprise Management are well documented. More and more companies have embraced Lean as the only way to manage their business. While the principles of waste reduction, human optimization, distributed design, and supply chain management has served many companies very well, these benefits may alternatively lead to absolute failure. The authors previously documented the “ravages” of lean thinking taken to the limit and proposed the Temporal Think Tank (T3™) [1] as a solution.

Lean management focuses on eliminating waste (non-value-adding activities) throughout their systems. This leads organizations to better understand their customers’ needs and deliver what the customer wants exactly when they want it (Just-in-Time). Lean has led to higher quality throughout the production chain and design through evolution. Major manufacturers concentrate their expertise and judiciously outsource vital sub-components and sub-assemblies to specialists who become critical to the supply of the final product. By selective sharing of design/development, the lead (lean) organization can focus resources and competencies on a limited set of innovative ideas and reduce their direct costs for innovating many of the components in their products. What, however, happens in the event of a sea-change as described in [1], or worse yet, the failure of a vital supplier? Only by addressing various scenarios can the high-level organization hope to succeed, or even survive.

This paper explores the positive and negative aspects of a lean philosophy. Providing this foundation exposes some pitfalls in lean thinking. The current downward shift in the marketplace poses important challenges to the lean paradigm. Can companies apply lean in a shrinking economy, while focusing on innovations needed for survival? In particular, this paper suggests that organizations, to protect their value streams, begin the effective use of business system- and production-oriented T3™’s, develop more intelligent and flexible manufacturing systems, redefine values to encourage long-term, risk-taking innovation activities, and closely support suppliers along their entire value chain.

* Corresponding author, R. Lindeke, Tel.: (01) 218-726-7947; Fax: (01) 218-726-7947; E-mail: rlindek1@d.umn.edu
2. BENEFITS AND PITFALLS OF LEAN SYSTEMS

While the benefits of lean practices are well documented, some of the negative effects of lean on employee outcomes, work characteristics, product design, and an organization’s innovation capability have also been studied [1-5]. To offer researchers and practitioners an extensive list of the potential pitfalls in the lean thinking, we present our analysis in this section. The analysis is based on a thorough literature review in the fields of lean manufacturing, new product development, technology and innovation management, and human resource management. Case studies are also employed in our research. Besides those ones identified in literature [1-5], many more issues have come to our attention as we link the lean philosophy to other management concepts. In the following sections, we attempt to clearly spell out the key concepts of lean and discuss the associated pitfalls.

2.1. REDUCING WASTE

Waste reduction is the fundamental concept in the lean philosophy. By focusing on value-adding activities, lean enterprises actively work to identify and eliminate different kinds of waste, called *muda* in Japanese, from the system. Any activities that consume resources but do not create value will be considered *muda*, and become the subject of Kaizen (continuous improvement) events. Wastes can include designing wrong products that no customer wants, making mistakes in the manufacturing process that leads to defects, employees being left idle or waiting for deliveries during work hours, etc. The seven categories of waste summarized in [6] have a good cover of all the wastes in lean definition, and they are listed as:

- Overproducing more items than included in customer orders.
- Inventory due to increases of finished goods and work-in-process.
- Motion that does not add value to the final product.
- Waiting for any resource throughout the flow of design and production.
- Transportation or the additional movement that is not of value to the product.
- Over-processing or additional steps that do not increase the overall value of the product.
- Not being right the first time or the costs and time associated with repairing and correcting a product.

Correct identification of *muda* and successful elimination of them leads to reduced manufacturing cost, higher product quality, improved customer satisfaction, and increased profits.

However, in reducing waste and focusing on value-adding activities, a company may focus on obvious short-term benefits and ignore long-term competitive advantages. Since the return on investment for many innovations is very difficult to quantify when the ideas first take shape, especially before the potential market has been clearly identified and developed, it is very likely that those ideas, especially the ones offering particularly long-term contributions, will be considered non-value-adding and thus be cut off.

Even before ideas get generated, the time and effort that are necessary to spark innovation may also be eliminated as waste in a lean system. It should be noted that when the lean concepts were first developed in Toyota Company by Taiichi Ohno, time was set aside periodically for the worker teams to get together and suggest ways to improve the system [7]. These “creative times” are highly valuable in keeping the teams, and hence the organization, innovative and to continuously improve the manufacturing process. However, as lean practices reach the extreme, the shrinking size of the workforce and the busy schedules of employees who have multiple responsibilities will make it much harder to get workers together for formal discussions, much less casual chats that may spark innovative changes.

In addition, in order to identify non-value-adding activities, value should be clearly understood and defined. In lean thinking, it is emphasized that value can only be defined by the end users [8]. Customers’ needs and wants should be followed closely in product design and manufacturing. Any products or features that the current customers do not want will be considered *muda* since they do not generate revenues in the current market. This approach is expected to continuously and incrementally improve products and increase customer satisfaction, especially in a market pull situation. However, it may hinder radical innovations that create technology push opportunities and cause companies to stumble upon disruptive innovations. A disruptive technology is a new technology or new application that usually underperforms in area(s) most desired by the mainstream customers, at least in the short-term, but with other valuable features and great potential to develop, the technology can eventually better more established technologies and dominate the market. As noted in [9], blindly following customers’ demand may lead a
company to focus on technology development that overshoots customers’ demand and lose the market to disruptive
technologies that had been initially denied by the same group of customers. Therefore, exclusively following
customers’ definition of value and the elimination of all “non-value” adding activities can lead a lean company to
failures because customers can be wrong, or at least short-sighted.

2.2 Human Resource Management

The second keystone to the lean philosophy is to “respect the workers” since they are the knowledge resources in
the system. As discussed above, slack in lean systems are continuously identified and removed as muda. Slack on
the human resource side mean unused work time and excess workers. Increasing worker utilization and reducing
the size of the workforce usually lead to reduced manufacturing costs. However, stress is also created in the work
environment from “crazy” schedules and multiple responsibilities. In certain situations, stress may drive “creative
tensions” that stimulate employees’ creativity [10]. However, too much stress is more likely to stifle employees’
creative thinking. As noted in [11], most people cannot function effectively in a time crunch for long periods
without burning out, even if they have a sense of being on an important mission and being challenged. Therefore,
even though the jobs in a mature lean system are supposed to make employees feel important and challenged and
thus greatly respected by the company, workers will not be able to innovate when they have been under too much
stress for too long. Typically when a company becomes leaner and leaner, a naturally expected result is layoffs.
Unlike workers in Japan, where lean manufacturing originated, workers in the West do not have the luxury of a
lifetime employment guarantee. As a result, lack of job security constitutes another source of employee stress in lean
enterprises. For employees in western lean companies, a conflict exists between the fear of losing jobs and the need
to accomplish more with less people while continuously eliminating excess workers. This probably explains why
critics renamed lean manufacturing “mean manufacturing” or “management by stress [3].”

Just because of being lean, the company runs the risk of being fragile in situations when many workers call in
sick or choose to go on strike. In either case, not enough people will be there to cover the jobs; work will be delayed
dramatically and quality may suffer. With “just enough” people who are very much stressed and in many cases
pushed to the limit, a lean system can hardly respond to such emergencies.

Employees in a lean company are trained and expected to take on multiple responsibilities. Since everyone can
perform every job to some degree, the flexibility of the system is increased. During times when sales decrease,
instead of waiting for orders beside the assembly line, primary engineers and assemblers are expected to transform
themselves into salespeople and go out to talk to potential customers to create sales. However, as a tradeoff,
employees may lose special expertise as they change roles frequently. In traditional organizations, the career ladder
is designed to allow employees to gain a depth of knowledge in one special area first and then expand their expertise
[12]. It is important to develop expertise in an area since it is essential to have expertise in order to innovate in such
an area, and broad understanding of different areas will be a plus that sometimes stimulates innovations. In lean
environments, if the job responsibilities are shared too broadly and shifted too often, the employees may never get a
chance to deepen their understanding and keep up with the development of technologies in any area. This will
eventually negatively affect the capability of an organization to innovate and can even lead to a regression in
retained organizational knowledge.

One of the cases that the authors considered when examining the potential pitfalls of a lean workforce is a
university department that has been short of teaching resources for several years. To survive increasing student
enrollment with a “lean” faculty group, many members have to take the responsibility of continuously teaching new
courses, some of which are not in their direct or closely related areas of expertise. At the same time, the average
teaching load and the service load have also been increased. Time to perform research and innovate, either with their
classes or in departmental outlook has been largely taken instead by developing course materials and preparing for
lectures. The innovation rate, measured by publications in this context, can only be seen to decrease.

2.3 Distributed Design

Distributed design characterizes the product design and manufacturing strategies in lean companies. In fact,
Japanese lean car producers design and provide detailed drawings of only 30% of the parts in their cars, the rest are
distributed to its first-tier suppliers, who usually have expertise in process engineering and plant operations[7]. In
this way, the responsibilities of manufacturing are mostly shifted to suppliers. The lean producers become focused
on the overall design of product and the final assembly of parts being engineered and produced by different
suppliers. This approach agrees with the economic theory that distributing jobs to the most efficient parties increases overall social welfare. By focusing on product design and final assembly while aggressively outsourcing parts, lean companies are able to steadily decrease the unit cost of products. However, since most major parts are outsourced as a “turnkey” project to suppliers, the lean producers are barely involved in manufacturing of many, and often key, components of their products. As a result, the company’s own ability to design, debug, and improve manufacturing systems, or even large segments of their products will decrease.

For a company that wants to achieve product leadership, it is critical to maintain and strengthen its manufacturing ability. This is linked to a third cornerstone of lean: being agile where this agility helps develop from the strength of design and manufacturing knowledge and facilitates bringing innovative products to market quickly. Designing high quality products and low cost manufacturing is an important part. When incremental changes to modules/parts no longer lead to overall product improvement, the relationships among the product components and core concepts may need to be reconfigured[13]. In such a situation, manufacturing processes for new designs will need to be created, assessed, and improved. The Toyota Company in Japan tries to maintain a long-term relationship with its suppliers and helps them improve manufacturing by loaning / switching its engineers through the supplier companies. This approach promotes communication and knowledge sharing between producers and suppliers to a great degree. However, to solve the fundamental problem, a company will need a strong incentive and great commitment of time and efforts to understand and be involved in major suppliers’ manufacturing process.

Another tradeoff to distributed design when parts are mostly outsourced is that the innovations from internal product research and development (R&D) at a company may flow out to competitors too quickly, even before the company can benefit fully. Since most suppliers provide parts to multiple customers, to reduce cost, suppliers are likely to provide parts that are similar, if not the same, to each customer. Even if there are agreements in place to keep trade secrets, other customers of the same supplier may “smell” innovations from the parts or assemblies they receive or conversations with the supplier’s representatives.

2.4 SUPPLY CHAIN MANAGEMENT

In the lean supply chain, parts from suppliers arrive shortly before they are needed since excess inventories were eliminated as waste. This approach saves storage space and costs, and improves efficiencies. However, with no reserve stocks, the system will stop running if parts do not arrive on time or a faulty shipment comes. This happened in the early stage when General Motors (GM) tried to embrace the lean philosophy in its Pontiac assembly plant in Pontiac, Michigan. Since all the buffers and inventories had been removed, a failure of a shipment from its supplier, in this case the Flint plant nearby, caused the entire plant in Pontiac to shut down and they had to send the workforce home four hours early with a loss in production of product that was slated to flow to the customer [8]. Although this issue is supposed to be solved by the supporting organizational structures that are constructed, there is no guarantee that a late or faulty shipment will not happen in a mature lean system. Supply problems become more likely when financial crises affect every component of the economy, as we find today. Since large complex parts of the product are “sole-sourced” in lean-supply systems, the failure of a supplier, especially a first-tier supplier, can be fatal to the company. Strikes at different levels of the supply chain will also have a significant negative effect on production at the lean company. One need only consider the problems at GM when a supplier of airbags went into bankruptcy and no longer provide products to them, or would even release GM’s tooling to move it to a different supplier [14]. The 1998 strikes in two of GM’s part supplier factories in Flint, Michigan, that caused 26 out of 29 North American assembly plants to close down and lay off nearly half of GM’s workforce [15] is another good example that without buffers, failure of any link on the lean supply chain will cause the whole chain to fall like dominos.

One of the prime driving ideas that most companies adopt when they begin their conversion to Lean management is to consolidate their vendor list. Since there is a cost associated with keeping communication channels open to vendors, even if they are not Tier 1 or even regular suppliers, most lean companies view extra vendors as muda. Reduction in vendors can save resources; however, it can also limit innovation. One of the authors spent time with a Lean manufacturer recently and observed an unfortunate issue as the engineering staff explored a new quality process. The engineers had identified a promising new measuring device but they were unable to purchase it for testing because the vendor was not “on the approved list!”. This led to a significant delay and additional expense for bringing innovation into the company. Surely this was an unexpected, but often observed, consequence of the effect of vendor and supply chain control in lean enterprises.
2.5 CUSTOMER MANAGEMENT

In the Japanese Toyota system, a large workforce is maintained to deal with customer management. Japanese buyers usually purchase a new car about every four years, due in large part to a governmental policy that requires very expensive and very demanding vehicle inspections as a car ages. In order to maximize the total income from a customer over the long-term, the sales department in Toyota devotes a significant, and costly, effort to keep good long-term relationships with every customer[7]. Keeping every customer’s information in a database and calling them frequently will instill customer loyalty and make it difficult for competitors to gain a share in the Japanese market.

Compared with customers in Japan, people in other countries do not have to purchase new cars as often in the absence of strict government regulations dealing with the inspection of aging vehicles. Therefore, maintaining a long-term relationship with customers seems much less important in those markets. A lean system would likely view a large workforce dedicated to customer management as a waste, thus, will train and expect the customer service personnel to perform other jobs. As mentioned in [7], during a visit to the Toyota company, the authors were told that the external relationship manager was too busy to meet with them since he was doing some assembling. In the same way, the experts that used to analyze and communicate customers’ needs may be moved to other jobs and become too busy with more obvious or urgent work, including assembly. By this action in the long-term, knowledge and expertise needed to managing customer interaction will be lost.

2.6 FINANCIAL SYSTEM

While the shadow pricing strategy is used in both lean and mass production companies, lean companies tend to use it more aggressively. As shown in Figure 1, lean companies try to set the selling price much lower than the actual initial cost, in hope that by applying lean practices, the unit manufacturing cost will drop faster and more dramatically as more customers are attracted to purchase the product. It is reasoned then that profits gained later through expanded market share will soon compensate for the losses incurred in the beginning of a product’s life. Increasing sales to the point that early losses can be compensated is therefore critical for the lean company to survive. While this strategy is expected to encourage initial purchases, applying it in a shrinking market without being able to differentiate the product with superior features and total qualities will only expedite the fall of the company. Simply being cheap is not the answer to sustaining success: innovation and technological advancements are the ultimate driving forces of product leadership.

In the Japanese lean system, long-term relationships exist between the lead company and its suppliers. Besides sharing personnel, they interlock equities and depend on each other for financing which in part justifies their reliance on shadow pricing as it spreads the wealth as well as the potential losses throughout the “extended” organization. Essentially, the lead company and its suppliers are sharing the same destinies. This requires the companies to be highly supportive to each other. Otherwise, failure of one member will quickly spread the crisis to the whole group.

![Figure 1: Shadow Pricing Strategy as Employed by Lean Organizations](image-url)
2.7 SUMMARY

Table 1 summarizes the lean concepts and the potential failures that they may lead to, as discussed in the previous sections.

Table 1: Summary of lean pitfalls

<table>
<thead>
<tr>
<th>Lean Concepts in:</th>
<th>Pitfalls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eliminate all wastes</td>
<td>Companies may:</td>
</tr>
<tr>
<td></td>
<td>• Eliminate the creative times that are necessary to innovations</td>
</tr>
<tr>
<td></td>
<td>• Focus on short-term value-adding activities and lose long-term competitive capability through radical innovations</td>
</tr>
<tr>
<td></td>
<td>• Miss technology-push opportunities</td>
</tr>
<tr>
<td></td>
<td>• Stumble over disruptive innovations</td>
</tr>
<tr>
<td></td>
<td>• Create more stresses in the work environment and make employees lose the feeling of job security</td>
</tr>
<tr>
<td>Multiple job responsibilities</td>
<td>Employees may lose expertise in their special areas</td>
</tr>
<tr>
<td>Distributed design</td>
<td>Companies may lose the ability to design, debug, and improve the manufacturing system of the parts</td>
</tr>
<tr>
<td></td>
<td>• Parts designed and manufactured in different places may not match</td>
</tr>
<tr>
<td></td>
<td>• Opportunities to develop radical and architectural innovations will be eliminated</td>
</tr>
<tr>
<td></td>
<td>• Internal innovations may flow out too fast before the company can benefit from its R&amp;D</td>
</tr>
<tr>
<td>Lean supply chain</td>
<td>• Late and faulty shipment of parts will stop the system</td>
</tr>
<tr>
<td></td>
<td>• The failure of a first-tier suppliers are fatal</td>
</tr>
<tr>
<td></td>
<td>• Effects of strikes on different levels of the supply chain will be amplified.</td>
</tr>
<tr>
<td>Customer management</td>
<td>Customer satisfaction may be lost in the long-term as a result of the busy schedules and multiple responsibilities of customer analyst</td>
</tr>
<tr>
<td>Financial management</td>
<td>• The use of shadow pricing strategy may lead to total failure in a shrinking economy</td>
</tr>
<tr>
<td></td>
<td>• Cross-ownership and interlocking equity with each other in the lean group may expedite the fall of the whole group in a crisis</td>
</tr>
</tbody>
</table>

3. SOLUTIONS

For a company to maximize the benefits of its lean practices, special attention must be paid to the pitfalls discussed above. While it is not the purpose of this paper to offer an exclusive list of solutions, a few are suggested.

When one reviews a company that has a strong survival instinct, often we find that they have embraced the ability to be agile and flexible in their product design and manufacturing environment. This agility is linked to their innovative spirit and is vested into the establishment of independent organizations, like those espoused by the author’s in the Temporal Think Tank™ (T3™) [1]. When economic conditions become difficult, as in the current worldwide economic crisis, it is the companies who have agile philosophies that can succeed. Thus, at least as it views the future, an organization must drop strict adherence to the Lean mantra of “its muda” to carve out its future by finding time to reflect and study for products that will allow it to flourish in the recovery period that is ahead. Creating an independent organization, like T3™ [1], with a cost structure not honed to achieve high profit margin in the current market is also a solution for companies to harness disruptive innovations without affecting the lean nature of the parent company.

To cultivate an innovation-encouraging culture in a lean company, it is important to define value in the way that innovation driving activities are viewed so as not to be considered muda. Setting aside some creative times not only helps reduce job anxiety and get employees re-energized, but also provides opportunities for the employees to interact and stimulate out-of-the-box thinking. Defining value effectively also calls for a positive attitude toward risk taking. As emphasized in an interview by Jeff Immelt, CEO of General Electric (GE), it is important for an organization to make it „ok“ to take risks and devote time and efforts to certain activities that do not produce results in a short term [16]. An R&D portfolio that builds both short-term and long-term competitiveness will better prepare
a lean company to stand out in fierce competition. As an essential part of the lean system, workers should be respected and involved as much as possible in the decision making process. In this way, a sense of importance, enjoyment, and satisfaction will be provided to motivate employees’ creativities.

Given the advantage of being flexible, a lean company should be able to respond to changes more effectively if contingency plans are developed ahead of their absolute need. Therefore, besides ensuring competitiveness through continuous innovations, the company also wants to closely monitor and forecast the market demand and technology trends, assess its own competitive advantages, conduct sensitivity analysis, and generate scenario analyses. A variety of decision analysis tools are available to facilitate such activities and sharpen decision makers’ judgment [17-20]. What the companies need to do is to select and adopt suitable ones to maximize their benefits. In addition, to better cope with uncertainties and disruptions in the dynamic market environment, more intelligent and flexible manufacturing systems are needed, especially for the companies in which tooling and equipment require large capital investment. Flexible product platforms can be designed to effectively share common components and deliberately project uncertainties into flexible elements that are carryover-modified in different product families or product generations. In this way, the need of redesign as well as changes in manufacturing tooling and equipment will be minimized in an event of new product introduction. Details about the systematic process of flexible product platform design can be found in [21].

Due to the strong dependencies of a lean company on its suppliers, it is critical for the lean company to support those suppliers. This calls for more transparent financial systems and cash flows along the chain, since the system is only as strong as the weakest component of the whole chain. In the unfortunate incident stated above, with the failure of a key supplier, one can only wonder if cost control tactics by GM were not ultimately to blame in the failure. Surely, squeezing a few more cents out of the cost of the product each year for the last few years, as its market share shrank, did little to improve the sales picture for GM but could have sealed the collapse of this supplier. With better foresight this cash strapped supplier, who was likely unable to support innovation, could have been saved and could have begun to deliver even better products to GM in the future. GM could have provided assistance, perhaps by increasing its prices by a few dollars per vehicle and passing the added revenue directly back to its struggling supply chain with a very small impact of its number of sales over the short term. This would be clearly seen as a “shadow pricing” initiative that would have reaped huge benefit into the future.

Any organization, regardless of where it lies along the road to lean, can open itself to clearly explain its financial condition to its partners. It can share difficulties and build long term trust, strength, and a healthful and supportive relationship, as suggested by the Japanese Lean Manufacturers, rather than creating legal adversaries as has happened here.

Closely working with the suppliers also helps the lean company to keep up with the most recent developments in manufacturing technologies. This may require additional personnel from the lean company to work in supplier companies; however, it is critical for the company to maintain its ability in designing and debugging the manufacturing system, and to optimize the whole value chain.

4. Conclusion

Lean enterprise management aims to eliminate waste, effectively manage personnel, distribute design among entities that are best at each stage, work with the supply chain, manage customers, and wisely manage the organization’s finances. In each activity, the focus is to eliminate muda, or waste.

For each of these elements of lean, examples have been provided in this paper to warn against the zealous over-application of lean. Taken beyond a moderate, reasonable level, lean will actually decrease the organization’s ability to compete. In times of great economic stress, implementing lean management as much as possible is an understandable reaction, but it may ultimately insure the failure of the organization rather than save it. The authors have offered advice to the newly lean organization to avoid many of the pitfalls of “over zealous lean” to become a stronger competitor for having survived the on going, or in fact any, financial crisis.
REFERENCES


A Methodology for Scheduling Part Families in Cellular Manufacturing Systems with Reconfigurable Machines

Ignacio Eguia\textsuperscript{1}, Jesus Racero, Sebastian Lozano, and Fernando Guerrero

\textsuperscript{1}Department of Industrial Organisation and Business Management
University of Seville
Seville, 41092, Spain

\textbf{ABSTRACT}

Cellular Manufacturing Systems (CMS) help to reduce set-up time, manufacturing lead-time, and improve productivity and operation control, but they lack in flexibility and, thus, cannot respond to the production of any quantity of highly customized and complex products. Some degree of adaptability and flexibility can be attained using reconfigurable machines in CMS. The aim of this work is to establish a methodology for an effective working of CMS in the presence of reconfigurable machines, which consists of some machine modules which can perform different operations. Products with similar characteristics must be grouped into families. Each family requires one configuration of machines and modules to produce all the products of the family. Set-up costs must be considered to change from one family to the next one. Machines and modules have been grouped into parallel cell configurations in order to reduce lead-time and improve productivity. A mixed integer linear programming model has been developed to select the product families and their configurations, to assign each family to each cell and to sequence the assigned families within the cells by minimum costs. The model has been tested with high size problems adapted from the literature.

1. INTRODUCTION.

Current manufacturing environment becomes turbulent and uncertain and some critical requirements for a manufacturing system are essential to survive. Short lead time, more variants, low and fluctuating volumes and low prices are some of these general requirements of next generation of manufacturing systems. On the basis of changes in the strategies to meet these requirements, different types of manufacturing systems have been developed, viz. the Dedicated Manufacturing Systems (DMS), the Flexible Manufacturing Systems (FMS), the Cellular Manufacturing Systems (CMS), etc. The DMS provides profitable and cost-effective production in a stable market, but it is unable to operate effectively in the present dynamic market scenario. The FMS use expensive CNC machines with fixed hardware and software to produce a variety of products, but their implementation has not been very successful because of the lower throughput, high cost or complex design.

A CMS is based on the group technology (GT) principle of grouping similar parts into families. In these manufacturing systems, machines are grouped into several cells, where each cell is dedicated to process one or more part-families. A CMS helps in reducing work-in-process, set-up time, manufacturing lead-time and material handling, and improve productivity. However, a CMS lacks in flexibility and, thus, cannot respond to some requirements such as dynamic part mix and demand variation. To reduce the level of performance deterioration under these dynamic manufacturing environments, some degree of adaptability can be attained using Reconfigurable Machines (RM) within a CMS. A RM is a modular machine with some basic and auxiliary machine modules. When the auxiliary modules are changed, different operations can be performed on the new machine configuration.

Usually a RM has been used as part of a Reconfigurable Manufacturing System (RMS). Koren et al. [1] defines a RMS as “a manufacturing system designed at the outset for rapid changes in structure as well as in hardware and software components in order quickly to adjust production capacity and functionality within a part family in response to sudden changes in market or in regulatory requirements.” Xiaobo et al. [2] considers a RMS as a
manufacturing system configured to produce a family of products that share some similarities. Time reduction for introducing new products to the market with high quality and low cost is a duty for enterprise survival in this new scenario. The manufacturing system must be able to yield different batch sizes from different product types, with the exact capacity and functionality in each case.

This research aims at the development of a methodology for an effective working of CMS in the presence of reconfigurable machines. To achieve this goal, the proposed methodology consists on three steps (figure 1): grouping products into families, selecting the appropriate cell for each family and scheduling the families for each cell. The requirements and needs for products in CMSs with reconfigurable machines are first discussed to group products with similar characteristics. Then, the development of a mathematical model is proposed to design and scheduling the best cost-effective cells. Finally, the main conclusions are presented.

The unique method in the literature applied specifically for CMS with reconfigurable machines is from Pattanaik et al. [3], who present a methodology to design machine cells using modular machines using certain characteristics of reconfigurable manufacturing. It is based on the clustering approaches used in Group Technology, to group machines into cells but in their research products are not grouped into families and modules are not assigned to the designed cells.

2. GROUPING METHODOLOGY.

In a CMS, machines are grouped into several cells, where each cell is dedicated to a particular part family and their objective is to maximize the cell independence. The identification of part families and machine groups is commonly referred to as cell formation. Numerous approaches have been reported for cell formation. A sequential procedure has been used in the present methodology to determine the part families first, followed by machine assignment.

The developed methodology for grouping products into families is based on hierarchical clustering agglomerative methods. The starting point for applying hierarchical clustering is to collect a data matrix where the
columns are the products to be cluster-analyzed and the rows of which are some attributes that describe the products. Reconfigurable machines are designed to possess a set of key characteristics, such as, modularity, scalability, integrability, convertibility and diagnosability [4]. In a previous research [5], some of these characteristics have been adapted to obtain the data matrix, and then calculating a resemblance matrix coefficient to measure the similarity among all pairs of products. Finally, using a clustering method, the values of the resemblance coefficient have been processed. As a result, a dendogram or inverted tree structure is obtained, that shows the hierarchy of similarities among all pairs of products, and the different grouping that can be formed depending on the similarity of products within a family. From the dendogram, the selection of families is made based on any criterion. Figure 2 shows an example of dendogram for four parts or products. There are four sets of families, indicated by four levels. The first one \((L=1)\) is composed of four families, each of which is composed of one product. The second one \((L=2)\) is composed of three families \((B-C), (A)\) and \((D)\), the next level \((L=3)\) is composed of families \((B-C)\) and \((A-D)\), and the last one is composed of one family.

![Dendogram for parts A, B, C and D results](image)

If the number of cells is a data, a trivial solution consists of selecting the level with exactly the same number of families. Then each cell is dedicated to produce all the products of each family. However, with reconfigurable machines the costs associated to change from a family to the next one are low and it may be “cheaper” assigning two or more families to one cell than only one family per cell.

The manufacturing system has to adapt its capacity and functionality to the production of each family. Therefore, in each system’s configuration, the capacity of the machines and the utilisation of their functionalities are parameters to optimise. A RMS tends to the usage of the full capacity and functionalities of the installed machines. The identified parameters have to be implemented with the same measurement: the cost. Thus, the key parameters for the product families’ selection and scheduling in RMSs are: the changeover (reconfiguration) cost, the cost of the under-utilisation of the machines’ capacity, and the cost of the under-utilisation of the machines’ functionalities [6].

This problem of selection and scheduling the product families at each level of the dendogram is quite similar to the Travelling Salesman Problem (TSP) which seeks to identify a Hamiltonian path (a tour) that minimise the distance travelled by the salesman. The goal of minimising total distance can be changed into minimising total cost or time. At first glance, several similarities with our problem arise. First, cities in TSP can be compared to product families in RMS. Second, the goal in TSP is to minimise the total distance/cost/time required for completing the tour, whereas in RMS the goal is to minimise the total cost. Finally, the salesman in TSP has to finish in the initial city, whereas a RMS is configured for producing the first family when the last one has finished. It is well known that the TSP is a combinatorial optimisation problem of NP class. Thus, the selection and scheduling of product families in RMS is a NP class problem too. In our case, with more than one cell, the problem of selecting the families, assigning to each cell and scheduling the selected families at each cell is then similar to the Multiple Travelling Salesman Problem (mTSP), which requires determining the cities to assign to “m” salesman as well as the optimal ordering of the cities within each salesman’s tour (figure 3). Thus, this problem is more complicated than the TSP.
In this paper, an integer programming formulation is proposed to assign, select and schedule part families in a cellular manufacturing system with reconfigurable machines based on a three-index formulation of Christofides et al. [7].

3. INTEGER PROGRAMMING FORMULATION.

For the selection and scheduling of the product families, a mathematical model that includes the cost parameters expressed above is required. This model is based on four assumptions: the capacities of the machines are considered infinite, there is only one process plan for producing each product, each machine has its own basic and auxiliary modules, and the number of cells is known. The notation used for model development (indices, parameters and variables) is shown in Table 1.

<table>
<thead>
<tr>
<th>Indices</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i, j )</td>
<td>Families</td>
</tr>
<tr>
<td>( l )</td>
<td>Level</td>
</tr>
<tr>
<td>( k )</td>
<td>Cell</td>
</tr>
</tbody>
</table>

| L | Number of levels |
| \( F_l \) | Set of families to produce in level \( l \) \( (l = 1, \ldots, L) \) |
| \( N_l \) | Number of families to produce in level \( l \), so \( N_l = |F_l| \) |
| \( R_{ijl} \) | Reconfiguration’s cost from family \( i \) to family \( j \) in level \( l \) \( (i \in F_l + \{0\}; \ j \in F_l + \{0\}; \ i \neq j; \ l = 1, \ldots, L-1) \) |
| \( H_0 \) | Cost of the under-utilisation of machines’ resources while producing family \( i \) in level \( l (i \in F_l) \) |
A Methodology for Scheduling Part Families in Cellular Manufacturing Systems with Reconfigurable Machines

\[ C \text{ Number of cells} \]

**Variables**

- \( T_{ijl} \): 1 if family \( i \) is produced just before family \( j \), within level \( l \) in cell \( k \) (\( i \in F_l + \{0\}; \ j \in F_l + \{0\}; \ i \neq j; \ l = 1, \ldots, L - C + 1; \ k = 1, \ldots, C \))
- \( K_l \): 1 if all the families within level \( l \) are produced (\( l = 1, \ldots, L - C + 1 \))
- \( U_{il} \): \( \geq 0 \), auxiliary variables for avoiding subtours (\( i \in F_l; \ l = 1, \ldots, \min(L - 2, L - C + 1) \))

The objective is to select the level of the dendrogram that minimises both the cost for changing between families (changeover costs), and the cost for the under utilisation of the resources of each machine while producing those families.

The constraints to take into account are:

1. Only one of the levels in the dendrogram is selected (constraint 0).
   - All the families from the selected level will be produced, and none of the families will be produced for the rest of the levels. Besides, the number of variables for the changeover of the production system (T) must be equal to the number of families (constraints 1).
2. In each level, the families are manufactured one by one following a production order, and at the end the system is configured for manufacturing the initial family. Variables T can exist before one family and after one family (constraints 2) only. The same occur in the TSP (Travelling Salesman Problem).
3. Each cell is used exactly once (constraints 3). In the model, an initial configuration \( \{0\} \) has been defined, so that the system starts and finishes at this initial state.
4. Subtours within each level are not feasible. Therefore, the Miller-Tucker-Zemlin constraints [8] are introduced. This set of constraints avoids that, as for example, five families were produced in two different ways, as: 1-2-3-1-… and 4-5-4-… (constraints 4).
5. Sign constraints of binary variables (T, and K) and auxiliary variables (U) (constraints 5.x).

Therefore, the mathematical model is the following:

\[
\text{Min } \sum_{l=1}^{L-C+1} \left( \sum_{i \in F_l + \{0\}} \sum_{j \in F_l + \{0\}} \sum_{k=1}^{C} R_{ijl} T_{ijl}^k \right) + \sum_{l=1}^{L-C+1} K_l H_{il}
\]

subject to

\[
\text{(0): } \sum_{l=1}^{L-C+1} K_l = 1
\]

\[
\text{(1): } \sum_{j \in F_l + \{0\}} \sum_{k=1}^{C} T_{ijl}^k = K_l \quad : l = 1, \ldots, L - C + 1; \quad j \in F_l
\]

\[
\text{(2): } \sum_{j \in F_l + \{0\}} \sum_{k=1}^{C} T_{ijl}^k - \sum_{j \in F_l + \{0\}} \sum_{k=1}^{C} T_{ijl}^k = 0 \quad : i \in F_l + \{0\}; \quad l = 1, \ldots, L - C + 1; \quad k = 1, \ldots, C
\]

\[
\text{(3): } \sum_{j \in F_l} \sum_{k=1}^{C} T_{ijl}^k = K_l \quad : k = 1, \ldots, C; \quad l = 1, \ldots, L - C + 1
\]

\[
\text{(4): } (N_l + 1) \cdot \sum_{k=1}^{C} T_{ijl}^k + U_{il} - U_{jl} \leq N_j \quad : i \in F_l; \quad j \in F_l; \quad j \neq i; \quad l = 1, \ldots, \min(L - 2, L - C + 1)
\]

\[
\text{(5.1): } T_{ijl}^k = \{0,1\} \quad : i \in F_l + \{0\}; \quad j \in F_l + \{0\}; \quad j \neq i; \quad l = 1, \ldots, L - C + 1; \quad k = 1, \ldots, C
\]

\[
\text{(5.2): } K_l = \{0,1\} \quad : l = 1, \ldots, L - C + 1
\]

\[
\text{(5.3): } U_{il} \geq 0 \quad : i \in F_l; \quad l = 1, \ldots, \min(L - 2, L - C + 1)
\]
As the number of families within level \( l \) (\( N_l \)) is \( L-l+1 \), then the number of variables and constraints may be calculated by the number of levels (or products) \( L \) and the number of cells \( C \).

Comments:
1. The selection of families can be solved calculating the cost of each level in the dendogram, and the level with the lowest cost will be selected. Thus, all the possible solutions are evaluated. A high number of products involve lots of calculations, which may require years for solving.
2. All the families are manufactured in sequential order, starting and ending with a predefined configuration.
3. The problem is highly complex (NP-Hard), and therefore the number of operations for the resolution of the model grow exponentially with the number of products. Therefore, the use of heuristic methods is suggested.
4. In the model, the capacity of the machines have been considered unlimited. Thus, the inclusion of the capacity of the machines and the product demands in future models are required.

4. EXPERIMENTAL RESULTS.

In order to solve the model, costs have to be calculated. As shown in Table 1, the model presents two different costs: the cost of the reconfiguration among families (R), and the cost of under-utilisation of the machine resources (H). Cost calculation is improved if it is decomposed in parameters. On one hand, the reconfiguration costs have been decomposed in three different parameters: \( \alpha_R \) if a module of machine is swapped, \( \delta_R \) if a machine is incorporated in the production system, and \( \gamma_R \) if a machine is removed from the system layout. On the other hand, the costs of under-utilisation of machine resources have been decomposed in two parameters: \( \alpha_N \) if a module of machine is not utilised, and \( \delta_N \) if a machine is not used when manufacturing a product. The costs assigned to those parameters are previously estimated. The estimation is an issue of great importance, because different costs involve the selection of different families.

Due to the lack of existing reconfigurable systems, some problems from the literature regarding Cellular Manufacturing Systems (CMSs) have been modified in order to be taken as CMS with reconfigurable machines problems. Data in CMS problems are machines and parts. Machine’s modules have been added. Table 2 shows the size of the different problems implemented regarding number of products and variables, and results as CPU time (in seconds) and Costs. Each problem has been solved with 1, 2 and 3 cells.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cell</td>
<td>2 Cells</td>
<td>3 Cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>8</td>
<td>57</td>
<td>0.09</td>
<td>248</td>
<td>28</td>
<td>0.28</td>
<td>483</td>
<td>19</td>
<td>0.08</td>
<td>702</td>
</tr>
<tr>
<td>Set 2</td>
<td>9</td>
<td>64</td>
<td>0.03</td>
<td>339</td>
<td>52</td>
<td>0.05</td>
<td>664</td>
<td>41</td>
<td>0.11</td>
<td>973</td>
</tr>
<tr>
<td>Set 3</td>
<td>10</td>
<td>111</td>
<td>0.23</td>
<td>450</td>
<td>93</td>
<td>0.09</td>
<td>885</td>
<td>75</td>
<td>0.28</td>
<td>1304</td>
</tr>
<tr>
<td>Set 4</td>
<td>11</td>
<td>112</td>
<td>1.08</td>
<td>583</td>
<td>86</td>
<td>1.67</td>
<td>1150</td>
<td>61</td>
<td>0.31</td>
<td>1701</td>
</tr>
<tr>
<td>Set 5</td>
<td>12</td>
<td>51</td>
<td>1.42</td>
<td>740</td>
<td>29</td>
<td>1.56</td>
<td>1463</td>
<td>17</td>
<td>1.94</td>
<td>2170</td>
</tr>
<tr>
<td>Set 6</td>
<td>13</td>
<td>117</td>
<td>0.06</td>
<td>923</td>
<td>96</td>
<td>0.09</td>
<td>1828</td>
<td>80</td>
<td>0.13</td>
<td>2717</td>
</tr>
<tr>
<td>Set 7</td>
<td>14</td>
<td>122</td>
<td>3.09</td>
<td>1134</td>
<td>89</td>
<td>1.95</td>
<td>2249</td>
<td>71</td>
<td>3.1</td>
<td>3348</td>
</tr>
<tr>
<td>Set 8</td>
<td>15</td>
<td>102</td>
<td>6.86</td>
<td>1375</td>
<td>75</td>
<td>13.89</td>
<td>2730</td>
<td>58</td>
<td>19.06</td>
<td>4069</td>
</tr>
<tr>
<td>Set 9</td>
<td>16</td>
<td>201</td>
<td>48.78</td>
<td>1648</td>
<td>159</td>
<td>42.25</td>
<td>3275</td>
<td>128</td>
<td>29.97</td>
<td>4886</td>
</tr>
<tr>
<td>Set 10</td>
<td>19</td>
<td>103</td>
<td>1.78</td>
<td>2679</td>
<td>91</td>
<td>6.06</td>
<td>5334</td>
<td>79</td>
<td>4.55</td>
<td>7973</td>
</tr>
<tr>
<td>Set 11</td>
<td>20</td>
<td>300</td>
<td>9.66</td>
<td>3100</td>
<td>266</td>
<td>5.58</td>
<td>6175</td>
<td>237</td>
<td>26.38</td>
<td>9234</td>
</tr>
<tr>
<td>Set 12</td>
<td>22</td>
<td>267</td>
<td>23.38</td>
<td>4070</td>
<td>214</td>
<td>284.45</td>
<td>9113</td>
<td>177</td>
<td>978.14</td>
<td>12140</td>
</tr>
<tr>
<td>Set 13</td>
<td>24</td>
<td>401</td>
<td>84.36</td>
<td>5224</td>
<td>366</td>
<td>380.22</td>
<td>10419</td>
<td>336</td>
<td>734.84</td>
<td>15598</td>
</tr>
<tr>
<td>Set 14</td>
<td>26</td>
<td>440</td>
<td>247.09</td>
<td>6578</td>
<td>389</td>
<td>772.81</td>
<td>13125</td>
<td>346</td>
<td>1679.95</td>
<td>19656</td>
</tr>
<tr>
<td>Set 15</td>
<td>30</td>
<td>433</td>
<td>285.34</td>
<td>9950</td>
<td>403</td>
<td>3316.8</td>
<td>19865</td>
<td>373</td>
<td>7034.41</td>
<td>29764</td>
</tr>
<tr>
<td>Set 16</td>
<td>37</td>
<td>659</td>
<td>2995.84</td>
<td>18315</td>
<td>587</td>
<td>1901.58</td>
<td>36588</td>
<td>547</td>
<td>12451.11</td>
<td>54845</td>
</tr>
</tbody>
</table>
The CPU time required for solving the proposed model is increasing with the number of products and with the number of cells. And the operational costs decrease with the number of cells. Those problems have been solved with the linear programming software CPLEX v.10, and using a CPU at 1.6 GHz PC. Figure 4 presents the variation of CPU time with the number of products for all the problems referred in Table 2.

In the figure 4, the tendency of CPU time with the number of products is increased hardly when the number of products is higher than 40 and using 3 cells. Therefore, the model is appropriate for problems with 40 products or less. This result shows that heuristics must be developed for solving problems with more than 40 different products.

5. CONCLUSIONS.

This research has presented a methodology to solve the cell formation and scheduling of part families in the presence of reconfigurable machines. The methodology uses agglomerative techniques of hierarchical clustering that were developed for the needs of the CMSs, conveniently modified in order to take into consideration the requirements of RMSs. For the selection of the product families, the assignment of those families to cells and the schedule in each cell, a mathematical model has been developed based on the identification of the key parameters that influence that selection. Finally, the model has been validated with 16 different problems from the literature, and their results have proved the need for developing heuristics when the problem presents more than 40 products and 3 cells.

ACKNOWLEDGEMENTS

This research has been partly funded by the Spanish Ministry of Science and Technology; project DPI2008-04788. The authors wish to acknowledge the Spanish Government for its financial support.
REFERENCES


Scheduling of production orders with assembly operations and alternatives

Krzysztof Kalinowski*

Institute of Engineering Processes Automation and Integrated Manufacturing Systems
Faculty of Mechanical Engineering, Silesian University of Technology
ul. Konarskiego 18A, 44-100 Gliwice, Poland

ABSTRACT
The problem of scheduling in flexible production systems is considered in this paper. The model of the scheduling system in which both machining processes and assembly operations are carried out, where there are alternative resources. Production order is described by multivariant technological process – a set of technological processes for the same product, which differ in technology, type of operations, treatments and operation times. Production orders with assembly operation are divided into single processes. The general scheduling procedure with searching best operation variant on the basis of the multivariant technological processes is also described. Time and cost based objective function is used for evaluation of a set of possible solutions. The proposed method can be used for searching acceptable solution, especially in cases, when it is not possible to find optimal solution.

1. INTRODUCTION
The increasing competition in the market and a balanced level of technology are forcing companies to use modern methods of organization of production processes. Adequate arrangements for the production flow lead to minimizing production costs, have a significant impact on the price of the product and allow the timely execution of orders. When creating a production schedule, it is necessary to use a data derived from both the organizational (dates, quantities – from MRP/ERP systems) and the technical and technological production preparation (structures of products, constructions and technology – CAD/CAM systems). Therefore, appropriate cooperation between these departments in an enterprise is required.

Before accepting a new order and taking actions connected with production, the main problem is to answer quickly the following question: Can a new order be realized according to the clients’ expectations, in the same quality and quantity, and a fixed period of time, with known constraints of resource consideration? Complexity of the problem of analysing all organisational variants makes it practically impossible to solve it in quantitative sense – to obtain the optimal solution in satisfactory time. For the same reason, it is also difficult to solve the problem of determining an acceptable production flow - that is, in qualitative sense.

Technological capabilities of manufacturing systems allow the creation of many different technological processes for the same item. The differences may relate to both technology implementations, the order of implementation of certain procedures (operations) and times (duration of the treatments, the setup times, etc.) related to the possibilities of resources, used tools and equipment. The technological process with more than one variant of the implementation in the production system is defined as the multivariant technological process. The current situation in the production system and a set of evaluation criteria in a particular situation should help in deciding on the selection of the best possible option for execution. The traditional approach to the design of technological processes with optimization of a fragmented approach, ignoring alternatives, significantly reduces the flexibility of its implementation. Multivariant technological processes development, under the conditions without computer support, may prove to be too time-consuming and unprofitable, especially when a large number of possible variants can be obtained. The CAPP systems (Computer Aided Processes Planning) do not have these restrictions.

* Corresponding author: Tel. +48 32 2372459; Fax. +48 32 2372459; E-mail: krzysztof.kalinowski@polsl.pl
Manufacturing systems with a large number of machines, in which both machining processes and assembly operations are carried out, where there are alternative resources, are commonly encountered in industry. For such systems, however, it is not possible to develop a schedule by using methods which guarantee obtaining the optimal solution. The existing solutions are based on the use of heuristic methods, dispatching rules, genetic algorithms, and hybrids.

2. THE MODEL OF PRODUCTION SCHEDULING

The discrete manufacturing system with concurrent, multi-assortment production is considered. The base of the production control is a schedule. Presented model satisfies the following basic assumptions:

- there is flexible job shop configuration of the production system,
- there is no limitation of work in progress stores capacity,
- the technological process of a product is considered as a multi-level configuration of machining and/or assembly operations,
- processes that are realised in the system may have more than one production route, determined by multivariant technological process.

Production system \( Ps \) is described by set of machines and set of production orders:

\[
Ps = (M, Po)
\]

where, \( M \) = set of machines, \( M = \{M_1, M_2, ..., M_h, ..., M_I\} \), \( I \) = number of machines; \( Po \) = set of production orders, \( Po = \{Po_1, Po_2, ..., Po_p, ..., Po_J\} \), \( J \) = number of orders.

Machine \( M_i \) is determined by:

\[
M_i = (C_i, a_i)
\]

where, \( C_i = \) calendar of availability periods of machine \( M_i \), \( a_i = \) type of machine.

Calendar gives the timetable including free periods of a machine for assigning operations. Type of machine is a feature describing machine ability for execution of group of specific operations. The production system contains \( h \) types of machine \((1 \leq h \leq I)\).

A production order describes a single product (or bath), and consists of one (in case of machining) or more processes (e.g. machining and assembly). Each assembly operation divides a production order between processes (subprocesses of an order). A production order has the following parameters:

\[
Po_j = (C_j, r_j, d_j, pr_j, b_j)
\]

where, \( C_j = \) set of precedence constrains of processes, \( r_j = \) release date, \( d_j = \) due date, \( pr_j = \) priority, \( b_j = \) batch size.

Set of processes precedence constrains defines relations between processes of a given production order. It can be described as:

\[
C_j = \{(P_{k1,j},P_{k2,j}), (P_{k2,j},P_{k3,j}), \ldots \}
\]

where, \( P_{k1,j} = \) process of production order \( P_j \) that preceded \( P_{k2,j} \), \( P_{k4,j} \) preceded \( P_{k3,j} \), \( k_1, k_2, k_3, k_4 = \) processes numbers, \((1 \leq k1 \leq k2 \leq k3 \leq k4 \leq Kj)\), \( Kj = \) amount of processes of order \( Po_j \).

An example of an order structure of a product with assembly operations is shown in figure 1. The table shows pre and post conditions of an order. The order consists of five processes. In the production system first the \( P_{3,1}, P_{4,1} \) and \( P_{5,1} \) should be executed, next \( P_{2,1} \) and next \( P_{1,1} \). First operations of \( P_{2,1} \) and \( P_{1,1} \) are assembly operations. In this case, set of processes precedence constrains is following:

\[
C_I = \{(P_{2,1,1},P_{1,1,1}), (P_{3,1,1},P_{2,1,1}), (P_{4,1,1},P_{2,1,1}), (P_{5,1,1},P_{1,1,1})\}
\]
Scheduling of production orders with assembly operations and alternatives

When multivariant technological processes are used, a process may have more than one variant of realisation (figure 2). It can be described as a graph:

\[ G_k = (V_k, D_k, \phi) \]  

where, \( G_k \) = \( k \)-th multivariant technological process, \( V_k \) = ordered set of machining states (nodes in the graph), \( D_k \) = set of all possible operations (edges in the graph), \( \phi \) = relations,

\[ V_k = \{S_{k1}, \ldots, S_{ks}, \ldots, S_{sk}\} \]  

A state of the product \((S_{sk})\) at any time of the production process is defined by a set of states of each elementary treatment:

\[ S_{sk} = (s_{k1}, s_{k2}, s_{k3}, \ldots, s_{kf}, \ldots, s_{kF}) \]  

where, \( S_{sk} \) = \( sk \)-th state of \( k \)-th process realization, \( s_{kf} \) – state of \( kf \) treatment, \( s_{k}\in \{0,1\} \), \( (kf = 1,\ldots,kF) \), \( F \) – amount of different states of process realization.

A state of the product description allows the identification of the state during the process implementation by identifying which of the elementary treatment has already been worked, and which are not. A set of all possible operations is described by:

\[ D_k = \{O_{i1,k}, O_{i2,k}, \ldots, O_{i3,k}\} \]  

where, \( O_{i2,k} \) = an operation of process \( P_i \) realized on machine \( M_{i2} \), \( i, i2, i3 \in \{1,\ldots, I\} \).

A process is represented by subset of operations:
Number of operations for realizing the process \( P_k \) must not be constant. It depends on selected path in the graph \( G \) (selected variants of operations). An operation variant is described by:

\[
O_{i,k} = (tb_{i,k}, t_{i,k}, ts_{i,k}, c_{i,k})
\]

where, \( tb_{i,k} = \) beginning time, \( t_{i,k} = \) operation time, \( ts_{i,k} = \) setup time, \( c_{i,k} = \) cost of operation variant.

3. SCHEDULING PROCEDURE

The procedure of creating a new schedule is shown in figure 3. The first step is to select a scheduling algorithm from a set of available algorithms. A scheduling algorithm is understood as a priority rule that organizes the processes according to a specific index. Each algorithm uses its own indicators, based on the parameters of the processes, their operations, deadlines, etc. Processes are then inserted into the schedule according to their priority and may be inserted according to the forward or backward scheduling method. In case of forward scheduling method the operations of a process are placed to a schedule sequentially from first to last, and the time they start is restricted by release time of the process or the completion time of previous operation. This method is used when a process should be made as soon as possible. The method of backward scheduling inserts operations of a process in order from last to first, and the time when an operation should be finished is restricted by due date of the process (last operation in a process) or beginning time of the next operation (the next operation in the technological process). This method is used most often when too early execution is not recommended (e.g. limited storages capacity, freezing of capital - losses due to storage of materials, reducing the period of usefulness, product deterioration).

Different variants of the operation are analysed in the next step of the scheduling procedure. The way of selecting the appropriate option is to schedule temporarily each possible operation variant and compare results. The best variant of operation (that determines the current production route) can be selected by the following criteria: \( \min \) cost, \( \min \) operation time, setup time, beginning/finishing time of operation. Generally, the objective is expressed by:

\[
\min_t \left( w_1 z_{i,k}^{tb} + w_2 z_{i,k}^t + w_3 z_{i,k}^{tr} + w_4 z_{i,k}^c \right)
\]

where, \( z_{i,k}^{tb}, z_{i,k}^t, z_{i,k}^{tr}, z_{i,k}^c = \) normalized coefficients of operation variant \( (O_{i,k} \neq \emptyset) \), \( w_1 - w_4 = \) weights.

Coefficient of starting time depends on the method of scheduling (equations 12 and 13):

\[
z_{i,k}^{tb} = \frac{tb_{i,k}}{\max_i \{tb_{i,k}\}}, \text{ in case of forward scheduling}
\]

\[
z_{i,k}^{tb} = \frac{\min_i \{tb_{i,k}^{sp} + t_{i,k}\}}{tb_{i,k}^{sp} + t_{i,k}}, \text{ in case of backward scheduling}
\]

After inserting all the operations of a process, the starting times (forward scheduling) or finishing times (backward scheduling) of associated processes are corrected. If all associated processes are scheduled, the next order is taken into consideration. Next, a complete schedule is evaluated using a set of criteria.
Scheduling of production orders with assembly operations and alternatives

Selection scheduling algorithm

Orders priority assignment

Selection of non-scheduled order with highest priority

Selection of scheduling method: forward or backward

Division of order to processes

Selection of non-scheduled process

Forward

Selection of non-scheduled operation from first to last

Backward

Selection of non-scheduled operation from last to first

Trial scheduling operation variants

Evaluation and selection of the best operation variant

Insert operation to the schedule

Are they inserted all the operations of the process?

Y

N

Correcting dates for starting and completion of related subprocesses

Are they inserted all the processes of the order?

Y

N

Are they inserted all the orders?

Y

N

Evaluation of generated schedule

Are they applied all the scheduling algorithms?

Y

N

Best schedule selection

Figure 3: The scheduling procedure
4. Multicriteria Evaluation

If as a result of scheduling is obtained more than one acceptable solution, it is necessary to carry out the evaluation and selection of solution for implementation. In real manufacturing systems there is more than one criterion that should be considered, so production schedule should be evaluated using set of criterions. To assess, taking into account more of the criteria, the application of multicriteria evaluation of the potential solutions is required. Method of determining the best variant of the schedule should take into account the value of time and cost based criterions for a specific decision-making situation.

In the proposed method of multicriteria evaluation of a schedule are the following steps:

- determining a set of criteria for evaluation,
- assigning the weights for each criterion,
- calculation of sub-ratings (for each criterion),
- calculation of the total assessment – aggregation of partial ratings.

The evaluation criteria may have different nature: deterministic, probabilistic or fuzzy. In the present work considers only deterministic criteria. A set of evaluation criteria, used in a given situation, is a subset of all available criteria (K):

\[ K \subset K_w \]

\[ K_w = \{K_1, K_2, ..., K_r, ..., K_R\} \]

where, \( K = \) set of criteria selected for evaluation, \( K_w = \) set of all available criteria, \( R = \) number of criteria.

The basic evaluation criteria adopted in the presented model are minimizations of: makespan (\( C_{\text{max}} \)), mean completion time (\( \bar{C} \)), maximum flow time (\( F_{\text{max}} \)), mean flow time (\( \bar{F} \)), maximum lateness (\( L_{\text{max}} \)), mean lateness (\( \bar{L} \)), maximum tardiness (\( T_{\text{max}} \)), mean tardiness (\( \bar{T} \)), maximum earliness (\( E_{\text{max}} \)), mean earliness (\( \bar{E} \)) and cost criteria.

For the selected evaluation criteria weighting coefficients (\( g_r \)) are established. Weighting coefficient determines the impact of a criterion for the final evaluation. After that, schedules are evaluated separately by each of the selected criteria, to determine the partial evaluations. The overall evaluation of a schedule is calculated by aggregating its partial evaluations. The best schedule is obtained from following objective:

\[ \min \sum g_r u_{r,e} \]

where, \( u_{r,e} = \) partial evaluation of the schedule \( H_e \) (\( 1 \leq e \leq E \)) according to the criterion \( K_r \), \( E = \) the number of created schedules.

5. Summary

The paper presents a model of a production system and the general procedure that enable creating an acceptable schedule. It can be used in most real production systems – as a solution at scheduling or new order verification stage. It is based on the idea of method of technological and organisational processes planning integration. The proposed solution is dedicated for enterprises that realise multi-assortment and changeable production. In enterprises, the validity of orders is generally known (orders have fixed priorities). This constraint allows abandoning the application of scheduling algorithms and focusing on searching the set of possible options for finding the best way of order realization. The development of the multivariant technological processes at the planning stage improves the process of scheduling and coordination of processes flow in the system in case of changeable conditions of production.

Determination of the solutions in changing external conditions, causing changes in the value of the input parameters, requires adequate organization of the scheduling system. An essential condition for generation of correct solutions in a short time is a quick access to required information. Therefore, the scheduling systems should not
operate separately, but must be integrated with databases of other systems (such as system monitoring, control, planning, management, etc.). In the present work, many important issues related to rescheduling are not taken into account, e.g. delays of schedule implementation that made it out of date even before its introduction. It is very important in designing and implementing a scheduling system in real manufacturing. It is forecasted in further researches.

ACKNOWLEDGEMENTS

This work has been conducted as a part of research project N115 081 31/1906 supported by Polish Committee of Scientific Research in years 2006-2009.

REFERENCES

Effect of Axial and Radial Velocities on Solids Mass Flow Rate Measurement

Jianyong Zhang*, Bin Zhou**

*School of science and Technology, University of Teesside, Middlesbrough, TS1 3BA United Kingdom
**School of Energy and Environment, Southeast University, Nanjing, China

ABSTRACT

In coal powder flow transportation and combustion, powder mass flow rate is a key parameter to be monitored and controlled. Electrostatics is one of the techniques used for such task with its non-intrusive, robust and low cost natures. The passive electrostatic meters measure charge induction on the detecting electrodes. As it is known that the induced signal is not only dependent on the solids mass flow rate, but also affected by solids velocity, however the velocity of particles referred to by far is the axial velocity. In reality, the solids velocity is a vector, its projections in both the radial and tangential directions also need to be investigated. This paper analyzes the dynamic sensitivity of ring-shaped electrostatic sensors using finite element method (FEM), and investigates the influence of the axial and radial velocities on the induced signal on the electrodes.

Keywords: electrostatic sensors, spatial sensitivity, velocity vector

1. INTRODUCTION

In pneumatically conveyed solids flow, electrification occurs due to friction, collision and separation between particles and between particles and pipe wall [1, 2]. Many factors, such as particle shape, particle size, coarseness of the inner surface of pipe wall, chemical composition of material, travel distance of particles, humidity, flow profile and solids velocity affect electrification process and thus charges carried by particles [3]. Although electrostatic charging creates problems for industries [4-7], it has also found various applications [8-11]. It is well known that the electrostatic technique has been employed to measure gas-solids flow [12-14], which is the subject to be discussed in this paper. The ring-shaped electrostatic sensors can be classified into two types according to probe (electrode) construction, the flush-mounted sensor and the lined sensor. A “flush-mounted” sensor, depicted in Fig.1 has its inner surface exposed to flow and is mounted flush with inner pipe wall [15]. This simple configuration offers high sensitivity to the charge carrying particles. In contrast, there is an insulator such as a ceramic tube inside a “lined sensor”, the electrode of such sensors has no direct physical contact with flow. In this paper, the analyses are based on the flush mounted sensors.

The axial velocity of charged particles has been measured using the cross correlation method [15] or the spatial filtering method [19] successfully. However the axial velocity provides only the mean velocity of the solids along the pipe axis. Consider any individual particle, or a small volume of air-solids mixture, its velocity is actually a vector due to the turbulent motion of particles. The velocity vector can be projected in the axial, radial and tangential directions in a column coordinate reference system. It has long been known that the solids velocity affects the charge induction on the electrode; obviously, it is inadequate to consider axial velocity only. The main purpose of the research presented in this paper is to establish the relationship between the velocity vector and the sensor’s output to study the sensing mechanism of the flush-mounted ring-shaped electrostatic sensor for air solids flow measurement. It is concluded that the dynamic sensitivity of a ring-shaped electrostatic sensor depends upon the amount of the charge carried by particles, the locations of charged particles, as well as on the velocities in both radial and axial directions.

2. Simulation based on 3-D FEM

2.1 Configuration of electrostatic sensor

Fig.2 (a) shows the configuration of a flush-mounted electrostatic sensor comprised of a short spool piece and a ring probe. The ring probe is surrounded by an insulator layer preventing it from conductive contact with the pipe wall as shown in Fig. 2 (b). The grounded metal pipe wall acts as an electrostatic screen naturally. The micro-column inside the pipe in the above figure represents a charged element, which can be regarded as a group of particles evenly dispersed in this column. The size of the element has to be small enough so that it can be regarded as a charge source surrounded by

* Corresponding author: Tel.: (0044) 1642-342546; Fax: (0044) 1642-342401; E-mail: j.zhang@Tees.ac.uk
the medium of uniform permittivity.

2.2 Mathematical model
The purpose of modeling is to find the relationship between the sensor’s output voltage and the amount of charge carried by an element, its location and its velocity. The modeling is based on electrostatic theory, and the magnetic field due to moving charged particles is ignored. The electrostatic equilibrium system is governed by Poisson equations under the given boundary conditions,

$$\nabla \cdot \varepsilon \varepsilon_0 \varepsilon (x, y, z) \nabla u(x, y, z) = \nabla \cdot D(x, y, z) = -\rho(x, y, z)
$$

$$u(x, y, z) \Big|_{(x, y, z) \in \Gamma_p} = 0$$

$$u(x, y, z, t) \Big|_{(x, y, z) \in \Gamma_e} = \text{const}$$

$$E_{\infty} = 0$$

where \(u(x, y, z)\) is the potential distribution. \(\Gamma_p\) represents the inner surface of the grounded pipe with zero potential, \(\Gamma_e\) is the floated ring probe, and its potential value can be regarded as a constant if a charge amplifier is connected to the electrode, which is commonly used for signal conditioning in electrostatic sensors. \(\rho(x, y, z)\) is the charge density distribution, \(\varepsilon_0\) and \(\varepsilon(x, y, z)\) are the permittivity in free space and relative permittivity distribution of particulate material respectively. \(D(x, y, z)\) is the electric flux distribution, \(E(x, y, z) = \varepsilon_0 \varepsilon (x, y, z) E(x, y, z)\). \(E(x, y, z)\) denotes the electric field strength, \(E_{\infty}\) is the electric field strength at the infinite point, \(E_{\infty}=0\). According to the electrostatic theory, if the distribution of relative permittivity \(\varepsilon(x, y, z)\) and charge density \(\rho(x, y, z)\) are known, the one and only one solution exists under the given boundary conditions to Equation (1). Thus, the distributions of \(D(x, y, z)\) and \(E(x, y, z)\) can be found, and the induced charge on the ring probe are given by,

$$Q = \int_S D(x, y, z) \cdot ds = \varepsilon_0 \int_S \varepsilon(x, y, z) \cdot E(x, y, z) \cdot ds$$

where \(S\) is the surface of the ring probe.

Consider the complexity of gas-solids flow, the flow pattern or particulate material distribution is capricious, the analytical solution to Equation (1) can hardly be found. The 3-D finite element analysis method is used to obtain a numerical solution instead.

For the 3-D finite element analysis, an equivalent integral form of Equation (1) can be expressed as follows, where \(F(u)\) is known as the “energy function”.

$$F(u) = \int_V \varepsilon \left( \frac{\partial u}{\partial x} \right)^2 + \left( \frac{\partial u}{\partial y} \right)^2 + \left( \frac{\partial u}{\partial z} \right)^2 - \rho u \right) \ dx \ dy \ dz = \min$$

$$u(x, y, z) \Big|_{(x, y, z) \in \Gamma_p} = 0$$

$$u(x, y, z, t) \Big|_{(x, y, z) \in \Gamma_e} = \text{const}$$

$$E_{\infty} = 0$$

The „Tetrahedron element“, depicted in Fig. 3, is used to discretize the whole field due to its simple coefficient matrix and adaptability to sophisticated field shape.

The relation between the potential anywhere inside the tetrahedron volume, \(\tilde{u}(x, y, z)\) and the potentials at its four vertexes, \(u_i\) \((i=1,2,3,4)\) can be expressed by the interpolation function:

$$\tilde{u}(x, y, z) = \sum_{i=1}^{4} N_i^e u_i$$

where \(N_i^e\) is the shape function.

After discretization and interpolation are processed, the total energy function \(F(u)\) of the volume can be obtained from that of individual element, and then the finite element equations can be constructed in the following way:

Dividing the entire volume of the system into \(e_0\) elements and \(N_0\) nodes, the total energy function \(F(u)\) can be expressed as...
$F(u) \approx F(\bar{u}) = \sum_{e=1}^{e_0} F_e(\bar{u})$ \quad (5)

where $F_e(u)$ is the energy function for each tetrahedron element.

Equation (3) can be solved by finding the extremum of $F(u)$. Equation (6) provides a necessary condition for $F(u)$ to be the extremum.

$$\frac{\partial F}{\partial u_i} = \sum_{e=1}^{e_0} \frac{\partial F_e}{\partial u_i} = 0 \quad (i = 1, 2, \cdots, N_0) \quad (6)$$

The discretization of equation (3) to each element results in,

$$F_e(u) \approx F_e(\bar{u}) = \int \left\{ \frac{\epsilon}{2} \left[ \left( \frac{\partial \bar{u}}{\partial x} \right)^2 + \left( \frac{\partial \bar{u}}{\partial y} \right)^2 + \left( \frac{\partial \bar{u}}{\partial z} \right)^2 \right] - \rho \bar{u} \right\} \, dx \, dy \, dz \quad (7)$$

where $V_e$ is the volume of an element.

Substitute $\bar{u}(x, y, z)$ with $\sum_{i=1}^{4} N_i^e u_i$ based on equation (4), the partial differentiation of $F_e(\bar{u})$ becomes,

$$\begin{bmatrix} \frac{\partial F_e}{\partial u_i} \end{bmatrix} = [K_e] [u] - [P_e] = 0 \quad (i = 1, 2, \cdots, n_0) \quad (8)$$

The coefficients matrix $[K]$, for each element is symmetrical. For the field divided into $e_0$ elements and $N_0$ nodes, matrices $[K], [P], [u]$, can be expanded into $[\tilde{K}], [\tilde{P}], [\tilde{u}]$, the total coefficient matrices for the whole volume, are:

$$[\tilde{K}] = \sum_{e=1}^{e_0} [K_e], \quad [\tilde{P}] = \sum_{e=1}^{e_0} [P_e] \quad (9)$$

From the above, the finite element equations can be expressed as

$$[\tilde{K}] [\tilde{u}] = [\tilde{P}] \quad (10)$$

For any given boundary conditions, the solution of equation (10) can be obtained.

The condition of charge equilibrium needs to be satisfied to the best in the numerical simulation, ideally the charge $Q$, carried by the charged element should be balanced by the sum of $Q_p$ and $Q_r$, the induced charge on the pipe wall and on the ring probe respectively. That is to say $(Q_p + Q_r)$ has the same magnitude as $Q_e$, but with opposite polarity. However, restricted by the computation resource, there have to be compromises between the calculation accuracy and the simulation length of the pipe. The length used in this paper is 6 times the internal radius of the pipe, $l=6R$, which can guarantee the accuracy without putting a heavy burden on the computation.

The finite element meshes for the sensor are depicted in Fig.4. Fig. (4) (a), (b) and (c) illustrate meshes on the pipe, ring probe and charged element respectively. Fig.5(a) shows that the local potential distribution is characterized spherically symmetrical with regard to the charged micro-column positioned in the geometry center of the pipe; (b) shows the distribution of electrical field strength on the inner pipe wall. The colour captions on the left top of (a, b) denote the levels of the potential and the electric field strength with blue as the lowest and red as the highest.

3. Analysis of Dynamic sensitivity

Generally, the ring electrode has to be connected to a preamplifier to amplify the induced signal and enhance its loading capability. Thus the frequency characteristics of measurement system are determined by the characteristics of the electrode and the conditioning circuit.

The “dynamic” electrostatic sensors detect the variation of the charge induced on the electrode. Approximately, the function of the conditioning circuit can be expressed as differentiation, i.e. the output voltage of the conditioning circuit is proportional to the rate of change of the induced charge,

$$U_o(t) \propto \frac{dQ(t)}{dt} \quad (12)$$

where $U_o(t)$ is the output of the conditioning circuit, $Q_i$ is the induced charge on the electrode, $\propto$ stands for “proportional to”.

Page 638
It can be seen that, for a given circuit, voltage $U_c(t)$ varies with $\frac{dQ_i}{dt}$, which depends on the location, radial and axial velocities of the moving charged micro-column element. The following derivation is to establish $Q_i$ and $\frac{dQ_i}{dt}$ as the function of the location $(x,r,\varphi)$ where a micro-column element with unity charge travels at the radial velocity $V_r$, axial velocity $V_x$, and produces $Q_i$ and $\frac{dQ_i}{dt}$. Hereafter $Q_i$ is expressed as $Q(x,r,\varphi)$ to indicate that it depends on the location $(x,r,\varphi)$ of the micro-column.

Due to the sensor’s symmetrical axial geometry, the angular position of the charged element does not affect the induced charge on the ring-shaped electrode. Hence $Q(x,r,\varphi)=Q(x,r)$. The simulation can be implemented in the rectangular region as depicted in Fig. 6. The origin $(x=0, r=0)$ is at the geometry center of the sensor.

For given radial positions, e.g. $r=0$ and 80mm for a sensor with the diameter of 14” (355.6mm) and the electrode width of 2mm, the induced charge, $Q(x,r)$ by the micro-column element of one coulomb charge originated from the location $(x=0, r=\text{constant})$ decreases with its axial location $x$ when this micro-column element moves in parallel with the pipe central axis $x$ towards either up or down stream direction. This can be seen in Fig. 7(a). Whilst for a given axial position, e.g. $x=0$ and $x=50$mm, $Q(x,r)$ increases monotonously with $r$, but it decreases with $r$ when the element approaching the edge of the sensing volume shown in Fig. 7(b).

The boundary of the sensing volume in this paper is defined as the edge where $Q_i$ has dropped to the value less than 1% of the maximum, the sensing volume under the above definition is shown in figure 8 where the belt represents the ring probe and the ellipsoidal body is the sensing volume. It is clear though that the sensing volume under this definition is much larger than the geometrical size of electrode along the pipe axis.

The following analytical expression via curve-fitting is derived based on the data contained in Fig. 7(a). From Fig. 9, it can be seen that a double Gaussian model offers high fitting accuracy

$$Q(x,r) = A(r)e^{-B(r)x^2} + C(r)e^{-D(r)x^2} \tag{13}$$

where $A,B,C,D$ are functions of $r$, the radial coordinate of the element for a given electrode.

Fig. 9 illustrates the variation of $Q(x,r)$ with $x$ at $r=80$, the analytical equation at this radial location is for the sensor with the dimensions given above.

$$Q(x,r) \mid_{r=80} = 0.2725e^{-1/3.793x^2} + 0.0953e^{-1/11.63x^2} \tag{14}$$

Figure 10 shows a charged element positioned at $(x,r,\varphi)$ on a circumference and the projections of its velocity vector $V_x$, $V_y$ and $V_z$ on the angular, axial, radial coordinates respectively. When a charged micro-column travels along this circumference, the induced charge on the ring probe won’t change due to the symmetrical configuration of the sensor. Thus the velocity $V_y$ does not contribute to the variation of the total induced charge; only $V_x$ and $V_z$ need to be considered.

Here the dynamic sensitivity $S_{\text{dynamic}}$ is defined as the derivative of induced charge on the probe, when the micro-column carrying unit coulomb of charges, moves at a velocity in the sensing volume:

$$S_{\text{dynamic}} = \frac{dQ_i(x,r,\varphi)}{dt} \tag{15}$$

Substituting $Q_i(x,r,\varphi)$ in Equation (13) with the right side of Equation (15), we have,

$$S_{\text{dynamic}}(x,r,V_x,V_z) = -2x(B(r)A(r)e^{-B(r)x^2} + D(r)C(r)e^{-D(r)x^2}) \cdot V_x + (A(r) \cdot e^{-B(r)x^2}) \cdot V_x + C(r) \cdot e^{-D(r)x^2} - A(r)x^2B(r) \cdot e^{-B(r)x^2} - C(r)x^2D(r) \cdot e^{-D(r)x^2}) \cdot V_x \tag{16}$$

It is clear from equation (16) that the dynamic sensitivity is dependant on the spatial position $(x,r)$, axial velocity $V_x$ and radial velocity $V_r$ of the charged element.

As an example, if this micro-column carrying the unit charge travels along the pipe line at $r=80$mm, the induced charge on the electrode of 355.6mm diameter and 2mm width, i.e. the dynamic sensitivity will be,
\[
S_{\text{dynamic}}(x, r, V_s, V_r) \big|_{r=80\text{mm}, V_y=0} = \frac{dQ(x, r)}{dt} = \left[-0.1437e^{-0.793x^2} - 0.0164e^{-1/11.63x^2}\right] \cdot x \cdot V_s \quad (17)
\]

This is a typical case of equation (16) when ignoring the radial velocity \(V_r\), as shown in Fig.11.

The effect of radial velocity on the sensitivity can be analyzed in the same way. According to the definition of the sensitivity in this paper, the output signal of the conditioning circuit is proportional to \(dQ / dt\) when the charged particle stream passes through the sensing zone at radius \(r\). For simplicity, assume the proportion factor is unity, the output of the conditioning circuit is

\[
\frac{dQ}{dt} = \sum_{(x, r, \varphi) \in \Omega} q(x, r, \varphi) \cdot S_{\text{dynamic}}(x, r, V_s, V_r) \quad (18)
\]

where \(q(x, r, \varphi)\) is the charged local element and \(S_{\text{dynamic}}(x, r, V_s, V_r)\) is the sensitivity, \(\Omega\) denotes the sensing volume.

4. Experiment and analysis

Fig. 12 shows a diagram of the belt rig in the laboratory of the University Teesside, which is used to validate the theoretical analysis discussed above. The rubber belt is pulled by two pulleys, one of which is driven by a DC motor. The speed of DC motor is adjustable to set the belt velocity from 5m/s to 25m/s. A brush is placed at one side to generate electrical charges. The charged belt is used to simulate a stream of charged micro-columns or charged particles traveling along axial direction. The diameter of the sensor is 14” (355.6mm), and the width of the electrode is 2mm. These data are used for numerical simulation in section 3.

Fig. 13 (a) shows the relationship between the belt velocity (\(V\)) and the root mean square value, \(\text{rms}\) of the sensor output at \(r=80\text{mm}\), the vertical coordinate is the normalized value of the output root mean square value. It can be seen that the value of \(\text{rms}\) increases with belt conveyance velocity. Fig. 13 (b) depicts the \(\text{rms}\) versus \(d\) at \(V_x=17\text{m/s}\), the distance of the belt to the pipe wall. It is clear the signal level decreases with \(d\).

5. Discussions

There may be three reasons for the signal increasing with the belt velocity. At first, higher velocity means more intensive friction between the belt and the brush, and between the belt and air, which produces more charges; secondly, even the charge carried by the belt unchanged, the sensitivity increases with axial velocity \(V_x\), it can be seen from Equation (16) and (17); in addition, if the radial movement of the belt represented by \(V_r\) is taken into account, the contribution from it may be more important.

It was observed that the belt vibration was more pronounced in the radial direction with the increased speed of the motor. The standard deviation to mean value ratio of \(\text{rms}\) was significantly higher than at lower velocity. In order to analyze the cause of the signal at the output of the meter under variable belt velocity, we replace the dynamic sensitivity in equation (18) with that in equation (16). It follows,

\[
\frac{dQ}{dt} = \sum_{(x, r, \varphi) \in \Omega} \left[-2q(x, r, \varphi) \cdot x \cdot (B(r)A(r)e^{-B(r\varphi)^2} + D(r)C(r)e^{-D(r\varphi)^2}) \cdot V_s \right. \\
+ q(x, r, \varphi) \cdot (A(r)e^{-B(r\varphi)^2} + C(r)e^{-D(r\varphi)^2} - A(r)x^2B(r)e^{-2(B(r)\varphi)^2} - C(r)x^2D(r)e^{-2(D(r)\varphi)^2}) \cdot V_s \}
\]

It can be seen that \(dQ/dt\) depends on \(q(x, r, \varphi)\) and its distribution, belt velocity in \(x\) and \(r\) directions.

Here presented is an interesting case in which the charges on the belt are assumed to be uniformly distributed over \(-l \leq x \leq l\), where \(-l\) and \(l\) are the boundaries of the sensing volume along a given flow line, from Fig.11, it can be seen that the sum of the terms associated with \(V_s\) in (19) approaches zero. Equation (19) can then be reduced into:

\[
\frac{dQ}{dt} = \sum_{(x, r, \varphi) \in \Omega} q(x, r, \varphi) \cdot (A(r)e^{-B(r\varphi)^2} + C(r)e^{-D(r\varphi)^2} - A(r)x^2B(r)e^{-2(B(r)\varphi)^2} - C(r)x^2D(r)e^{-2(D(r)\varphi)^2}) \cdot V_s \\
= \sum_{(x, r, \varphi) \in \Omega} q(x, r, \varphi) \cdot K(r, x) \cdot V_s
\]

\[
(20)
\]
where \( K(r,x) = A(r) \cdot e^{-B(r)x^2} + C'(r) \cdot e^{-D(r)x^2} - A(r)x^2 B'(r) \cdot e^{-B(r)x^2} - C(r)x^2 D'(r) \cdot e^{-D(r)x^2} \)

which is independent of \( q(x,r,\phi) \) and \( V_r \). In this sense, the output signal of the sensor is related to the radial velocity and the amount of charge carried by the moving belt. The velocity \( V_x \) does not affect the signal.

It cannot be verified in reality if the charges were uniformly distributed along the belt or not. However, for the rig with a belt of uniform property it is reasonable to imagine that there could be more contribution from the radial movement to the sensor’s output than that from the movement along the pipe axis.

In pneumatic conveying system, it has also been observed that at high velocity, the air solids flow is more turbulent, and the signal rms value is higher. This is due to the fact that there is intensified solids radial movement. This may also confirm that the effect of the radial movement of the solids is dominant.

6. Conclusions

In this paper, the 3-D finite element method (FEM) is employed to investigate the dynamic sensitivity of the electrostatic sensor with a ring-shaped probe mounted flush with the inner surface of the pipe wall. Theoretical analyses reveal the dynamic sensitivity depends on the spatial location, the axial and radial velocities of the charged element. From experiments using a belt rig, it confirmed the theoretical analysis regarding the relation between sensitivity and axial velocity as well as the relation between sensitivity and flow stream spatial position. This is the first attempt to identify the effect of the radial velocity on the output signal of the dynamic electrostatic sensor. The test results seem to support the theoretical analysis in this aspect too, although further in depth investigation is required.

Acknowledgement

This work was funded by European Research Fund for Coal and Steel, the contract number RFC-CR-03005.

Reference:

Fig. 1 Flush-mounted Sensor

Fig. 2 (a) Configuration of electrostatic sensor

Fig. 2 (b) Ring probe on insulator

Fig. 3 Tetrahedron element

Fig. 4 Meshes on pipe, probe and Charged element

Fig. 5(a) the local potential distribution

Fig. 5(b) Electrical Field strength distribution

Fig. 6 Simulation region

Fig. 8 Contour of sensitive region
Fig. 7 (b) Variation of $Q/r$ with $r$

Fig. 9 Double Gaussian curve fitting

Fig. 10 Projections of velocity

Fig. 11 Dynamic sensitivity variation along a flow line ($r=80$mm)

Fig. 12 Conveyance belt rig
Fig. 13 Relationship between the belt velocity and the root mean square value
An Efficient Algorithm Based on TOC to Solve Product Mix Problems

Alireza Rashidi Komijan\textsuperscript{1} and Saloomeh Yazdi\textsuperscript{2}

\textsuperscript{1}Department of Industrial Engineering
Islamic Azad University, Firoozkuh Branch
Firoozkuh, Iran

\textsuperscript{2}Department of Planning and Development
Production Research Center
Tehran, Iran

\textbf{ABSTRACT}

Product mix problem (PMP) is one of the most important and complicated problems in production systems. Proper determination of product mix leads to high level of profit because it specifies the best way of assigning work centers capacity to each product. In other words, it clears how many of each product should be produced in order to reach a high level of profit. Different approaches have been applied to solve this problem, among them, theory of constraints (TOC) has been widely considered since 1990s. This paper develops a distinguished algorithm to solve product mix problems that is efficient both in single and multi-bottleneck problems. At first, the proposed algorithm uses a mathematical model to aggregate different priorities assigned to products by different bottlenecks and finds an initial solution. Then tries to improve the solution by solving a set of linear inequalities. In this paper, it is discussed that the proposed approach reaches better solutions than the previous algorithms.

1. \textbf{INTRODUCTION}

Theory of constraints (TOC) is a production planning philosophy that was introduced in The Goal [1]. It aims to improve the system throughput by efficient use of bottleneck(s). When there is a bottleneck in the system, the demands of all products cannot be fully met. In this case, the problem is to determine product mix in such a way that high level of profit is obtained. This is called the product mix problem. Although PMP can be formulated as an Integer Linear Programming (ILP) model, it cannot be solved easily in large scale problems. If the number of products increases, the difficulty of solving ILP model will increase exponentially. As a result, many researchers focused on the heuristic algorithms.

Some of the researchers developed algorithms to solve PMP in single bottleneck environment whereas others focused on multi-bottleneck systems. Although many algorithms have been developed to solve PMP in multi-bottleneck systems, most of them fail in reaching desirable solutions. The main reason of their failure is that all decisions are made based on the most capacitated constraint (dominant bottleneck). The common process in the existing algorithms is that the product mix is determined based on priorities of products, i.e. the product with the higher priority is produced first. These priorities are assigned by bottlenecks to products. Each bottleneck assigns a priority (weight) to each product. Most of the previous researches just used the priorities assigned by the dominant bottleneck in determining product mix. Considering only one bottleneck in decision making and ignoring the importance of other ones may distort the solution because only a part of information is used. It is similar to searching for the local optimum instead of the global one.

In this paper, all bottlenecks are considered in product mix decision simultaneously. Each bottleneck assigns a weight to each product. So, there will be \( k \) weight values for each product (\( k \) is the number of bottlenecks). A mathematical model is used to aggregate these values and an aggregated weight is assigned to each product. The initial product mix is obtained using the aggregated weights. Then, the possibility of improving solution is examined through "decrease and increase process". In this process, it is tried to improve the profit by decreasing the

* Corresponding author: Tel.: (98) 912-3242816; Fax: (98) 21-88811727; E-mail: alireza_rashidi@yahoo.com
production amount of one product and increasing the other one. In the proposed algorithm, alternatives to decrease and increase are defined and the best one is selected easily. This causes the speed of solving a problem by the proposed algorithm to be independent from the dimension of the problem. So, the proposed algorithm can be applied in large scale problems properly. In the proposed algorithm, decrease and increase process is done simply by solving a set of linear inequalities. Besides the simplicity of the proposed algorithm, it is shown that it will reach better solution than the existing algorithms. In Sections 3 and 4, the proposed algorithm and its logic are described in details.

2. LITERATURE REVIEW

Solving PMP using TOC heuristic is an interesting area for many researchers. Several algorithms have been developed on this topic since 1990. The first (traditional) algorithm based on TOC was verified by several researchers [2,3]. They showed that the traditional algorithm could lead the optimum solution as ILP.

Balakrishnan and Cheng [4] criticized the traditional algorithm and showed that it could not reach optimum solution in multi-bottleneck systems. Lee and Plenert [5] showed that the algorithm was inefficient in handling problems in which new product alternatives were to be added to an existing production line. Plenert [6] showed by an example that the traditional algorithm was inefficient in utilizing bottleneck resource properly.

Fredendall and Lea [7] revised the traditional algorithm using the concept of dominant bottleneck. The dominant bottleneck is the most capacitated resource. In the revised algorithm, dominant bottleneck assigns a weight (priority) to each product and the initial product mix is developed based on these weight values. Then, the revised algorithm tries to improve the solution using the neighborhood search concept. Fredendall and Lea [7] claimed that the revised algorithm could reach the optimum solution in all cases.

Aryanezhad and Rashidi Komijan [8] discussed several disadvantages of the revised algorithm and showed that it could not reach the optimum solution in all cases and might be stopped in a non-optimum solution. They also showed that the revised algorithm faced burdensome computations when the number of products increases. They discussed that stop condition of the revised algorithm was not defined properly as it might cause the algorithm to reach a non-optimum solution. Then they developed their own algorithm which is called "improved algorithm". It starts with an initial solution and then finds the best path to reach the improved solution with a logical procedure. The ability of improving solution by the improved algorithm is much more than the revised one but however, it is also limited. Tsai and Lai [9] expanded the improved algorithm to systems including joint products.

Hsu and Chung [10] developed an algorithm that was similar to the algebraic concept of simplex to some extent. The main assumption of this algorithm is that the model is continuous not integer. Clearly, there is no need to heuristic algorithms when the problem is considered continuous as simplex can solve large scale problems using advanced softwares like GAMS.

Souren et al. [11] categorized conditions in which TOC heuristic can reach optimum solution and discussed problems that it fails in reaching optimum.

Rashidi Komijan and Sadjadi [12] considered throughput and late delivery cost as decision making criteria and solved PMP problem using TOPSIS technique.

Among other researches, we can mention in [13], [14] and [15]. Kee and Schmidt [13] developed a general model including TOC and Activity-Based Costing (ABC). Bhattacharya and Vasant [14] solved PMP in such a way that decision maker reached higher degree of satisfaction and lesser degree of fuzziness.

3. THE PROPOSED ALGORITHM LOGIC DESCRIPTION

In this section, the logic of the proposed algorithm is discussed in details. In almost all PMP algorithms, the production plan is developed based on priorities of products. These priorities (weights) are determined by the bottlenecks. If there is one bottleneck in the system, there will be a unique weight vector and production plan can be easily obtained. As the resulted product mix is optimal in view of the bottleneck, it is optimal for the whole system. In a multi-bottleneck system, each bottleneck presents its own weight vector and as a result, there may be several weight vectors. The question is that which one of the weight vectors should be considered in developing production plan. Some preliminary algorithms used the weight vector associated to the dominant bottleneck. Naturally the
result may be good in view of that bottleneck not necessarily the others. Considering only one bottleneck in determining product mix and ignoring the importance of other ones is a pitfall for an algorithm.

The proposed algorithm overcomes this problem letting all bottlenecks contribute in determining product mix. In the proposed algorithm, bottlenecks and products are treated as decision-makers (DMs) and alternatives respectively. Using this analogy, the product mix problem is considered as a group decision making problem. Each DM assigns a weight to each alternative (individual weights). Also, as DMs (bottlenecks) do not have equal importance, the weights of DMs are determined in the proposed algorithm. Then, individual weights are aggregated through a mathematical model. The initial production plan is developed using the aggregated weights.

The rest of the algorithm is dedicated to improve the initial plan. This is done using "decrease and increase process". Through this process, the scheduler may decrease the production amount of a product and increase the other one provided that the process leads to a higher profit. Clearly there are too many alternatives for decrease and increase even in a medium scale problem. There are some features in the proposed algorithm that cut the number of alternatives considerably. In the algorithm, it is discussed that the product candidate to decrease should have higher priority that the product considered to increase. On the other hand, as the initial plan is based on the aggregated weights, it has minimum derivation from the desired plan. So the number of alternatives for decrease and increase process is limited and the time consumed in reaching the solution is decreased considerably. Decrease and increase process is applied in [7] and [8] but in the algorithm, it is done in a completely different manner that leads to better solution than the previous algorithms. The proposed algorithm is described in Section 4 and through a numerical example, its result is compared to traditional, revised and improved algorithms. It is shown that the proposed algorithm is more efficient than the previous ones.

4. THE PROPOSED ALGORITHM

The following notations are used in the proposed algorithm:

- $i$: Index of product, $i=1,...,n$.
- $j$: Index of resource, $j=1,...,m$.
- $t_{ij}$: Processing time of product $i$ on resource $j$.
- $D_i$: Demand of product $i$.
- $P_i$: Produced units of product $i$.
- $SP_i$: Selling price of product $i$.
- $RM_i$: Raw material cost of product $i$.
- $CM_i$: Contribution margin of product $i$.
- $AC_j$: Available capacity of resource $j$.
- $RC_j$: Required capacity of resource $j$.
- $PR_{ij}$: Priority of product $i$ assigned by bottleneck $j$.
- $F_{ij}$: Weight of product $i$ assigned by bottleneck $j$.
- $W_i$: Aggregated weight of product $i$.
- $W'_j$: Weight of bottleneck $j$.
- $L_j$: Time left in bottleneck $j$.
- $H_j$: Maximum number of iterations according to bottleneck $j$.
- $H$: Real maximum number of iterations (according to the group of bottlenecks).

Step 1. Identify the system bottleneck(s): Each resource that its required capacity exceeds the available one, is considered as a bottleneck. Required capacity of resource $j$ is calculated as follows:
If $RC_j > AC_j$, then $j$ is a bottleneck.

Step 2. Determine the weight of each product assigned by each bottleneck (individual weights): For each bottleneck, different products have different priorities. Priority of product $i$ in view of bottleneck $j$ is calculated as follows:

$$PR_{ij} = \frac{CM_i}{t_{ij}}$$  \hspace{1cm} (2)

where $CM_i$ is calculated as follows

$$CM_i = SP_i - RM_i$$  \hspace{1cm} (3)

Weight of product $i$ in view of bottleneck $j$ is calculated by normalizing $PR_{ij}$ values:

$$F_{ij} = \frac{PR_{ij}}{\sum_{j=1}^{n} PR_{ij}}$$  \hspace{1cm} (4)

Typically, $(F_{ij} ,..., F_{ij} ,..., F_{ij})$ is the weight vector of products in view of bottleneck $j$. Each bottleneck has a weight vector expressing the weights that it assigns to different products. These weight vectors are called individual weights.

Step 3. Develop production plans regarding the weights obtained in Step 2. In this step, $m$ production plans are developed based on the weights calculated in Step 2. In other words, each bottleneck develops the production plan regarding its own weight vector. Clearly, these plans should be feasible, i.e., the production plan developed by one bottleneck should be feasible for others.

Step 4. Determine the weights of bottlenecks. To determine the weights of bottlenecks, the profit related to each production plan should be calculated. Normalizing these profit values will be led to the weights of bottlenecks.

Step 5. Determine the weight of each product in view of the group of bottlenecks (aggregated weight). In order to calculate the aggregated weight of a product based on the individual weights obtained in Step 2, a mathematical model is formulated as follows:

$$\text{Min } \sum_{j=1}^{m} \sum_{i=1}^{n} W'_{ij} |W_i - F_{ij}|$$  \hspace{1cm} (5)

Subject to:

$$\sum_{i=1}^{n} W_i = 1$$  \hspace{1cm} (6)

$$\min\limits_{j} (F_{ij}) \leq W_i \leq \max\limits_{j} (F_{ij}) \quad i = 1,2,...,n$$  \hspace{1cm} (7)

$$W_i \geq 0$$  \hspace{1cm} (8)
Step 6. Develop production plan regarding the aggregated weights. In this step, the production plan is developed by the group of bottlenecks using the aggregated weights obtained in Step 5. The profit resulted from the aggregated plan may be improved. The improvement will be occurred during a tradeoff, i.e. when the produced amount of one product is decreased and the time left in bottlenecks is dedicated to another product. In the following steps, the possibility of improving the profit is examined.

Step 7. Determine allowable tradeoffs. Decreasing production amount of product $i$ and increasing product $l$ is considered as an allowable tradeoff if $W_i \geq W_l$. In other words, if $W_i < W_l$, decreasing $i$ and increasing $l$ is useless. The reason is simple: In the initial plan, at first we considered $l$ for production and allocated bottlenecks time to it and produced $P_l$ units. If it was possible to produce additional units of $l$, it would be done in the initial plan automatically. Without regarding to the mathematical model, there may be $2!/(n-2)!$ alternatives for decrease and increase process. Using the mathematical model and the condition described in Step 7, the number of alternatives will be cut significantly.

Step 8. Determine how many $l$ should be increased in turn of decreasing one unit of $i$. Assume that $k$ is the number of product $l$ that should be increased in turn of decreasing one unit of $i$. $k$ is the least integer that satisfies the following inequality:

$$k \cdot CM_j \geq CM_i$$  \hspace{1cm} (9)$$

So, in each iteration, it is allowable to decrease one unit of $i$ provided that $k$ units are added to $l$. Now the question is that how many iterations are allowable to be done. This is answered in the following step.

Step 9. Determine the maximum number of iterations. Assuming that decreasing one unit of $i$ and increasing $k$ units to $l$ as one iteration, we want to know that how many iterations are allowable to be done. $L_j$ is the time left in bottleneck $j$ regarding production plan based on aggregated weights and before starting decreasing and increasing process. Decreasing one unit of $i$ causes that $t_{ij}$ minutes are added to the time left in bottleneck $j$ and increasing $k$ units to $l$ consumes $k t_{lj}$ minutes. If $k t_{lj} - t_{ij} \leq 0$, the iterations can be done until production units of $i$ falls to zero or the demand of $l$ is fully met. If $k t_{lj} - t_{ij} > 0$, the iterations can be done until production units of $i$ falls to zero, the demand of $l$ is fully met or bottleneck $j$ exhausts. These can be summarized as follows:

If $k t_{lj} - t_{ij} \leq 0$

Then $$H_j = Min (P_i, \left\lfloor \frac{D_l - P_l}{k} \right\rfloor)$$

Otherwise $$H_j = Min (P_i, \left\lfloor \frac{D_l - P_l}{k} \right\rfloor, \left\lfloor \frac{L_j}{k t_{lj} - t_{ij}} \right\rfloor)$$

where $\left\lfloor \frac{D_l - P_l}{k} \right\rfloor$ and $\left\lfloor \frac{L_j}{k t_{lj} - t_{ij}} \right\rfloor$ are integer values of $\frac{D_l - P_l}{k}$ and $\frac{L_j}{k t_{lj} - t_{ij}}$ respectively.

As the tradeoff must be feasible for all bottlenecks, the above calculation is repeated for all bottlenecks. Each bottleneck suggests a certain number of iteration ($H_j$) and the real maximum number of iterations is the minimum of values suggested by the bottlenecks:

$$H = \min_j H_j$$  \hspace{1cm} (10)$$
Step 10. Calculate net profit change related to the tradeoff between $i$ and $l$. In the previous steps, it was cleared that:

a) It is allowable to decrease $i$ and increase $l$.

b) $k$ units of $l$ should be increased in turn of decreasing one unit of $i$.

c) Part (b) can be repeated for $H$ times.

Clearly, net profit change is calculated as follows:

$$\text{Net Profit Change} = H (k \cdot CM_l - CM_i)$$  \hspace{1cm} (11)

Steps 8 to 10 is repeated for all allowable tradeoffs determined in Step 7. The tradeoff with the highest net profit change is selected.

5. NUMERICAL EXAMPLE

The proposed algorithm is explained through an example. This is the example that was discussed by Hsu and Chung [7]. Assume that a factory produces four products: $R$, $S$, $T$ and $U$. Demand, selling price and raw material cost of the products are shown in Table 1. The contribution margin (CM) is the difference between the selling price and the raw material cost. The factory uses seven resources: A, B, C, D, E, F and G. The processing time of each product in each station is presented in Table 2.

| Table 1. Weekly demand, selling price and raw material cost of each product |
|---|---|---|---|---|
| **Product** | **Demand (Unit)** | **Selling Price ($)** | **Raw Material Cost ($)** | **CM ($)** |
| R | 70 | 100 | 20 | 80 |
| S | 60 | 120 | 60 | 60 |
| T | 50 | 110 | 60 | 50 |
| U | 150 | 50 | 20 | 30 |

| Table 2. Processing time in minute, available and required capacity |
|---|---|---|---|---|---|---|---|
| **Resource** | **Product** | **A** | **B** | **C** | **D** | **E** | **F** | **G** |
| **R** | 20 | 5 | 10 | 0 | 5 | 5 | 20 |
| **S** | 10 | 10 | 5 | 30 | 5 | 5 | 5 |
| **T** | 10 | 5 | 10 | 15 | 20 | 5 | 10 |
| **U** | 5 | 15 | 10 | 5 | 5 | 15 | 0 |
| **Available Capacity** | 2400 | 2400 | 2400 | 2400 | 2400 | 2400 | 2400 |
| **Required Capacity** | 3250 | 3450 | 3000 | 3300 | 2400 | 3150 | 2200 |
| **Difference** | -850 | -1050 | -600 | -900 | 0 | -750 | 200 |

Step 1. According to the last row of Table 2, A, B, C, D and F are bottlenecks.

Step 2. Using contribution margin of each product (the last column of Table 1) and processing times in Table 2, the products priorities in view of each bottleneck are obtained:

| Table 3. Products priorities in view of each bottleneck ($P_{R_i}$) |
|---|---|---|---|
| **Bottleneck A** | 4 | 6 | 5 | 6 |
| **Bottleneck B** | 16 | 6 | 10 | 2 |
| **Bottleneck C** | 8 | 12 | 5 | 3 |
| **Bottleneck D** | 0 | 2 | 3.33 | 6 |
| **Bottleneck F** | 16 | 12 | 10 | 2 |
An Efficient Algorithm Based on TOC to Solve Product Mix Problems

Normalizing each row will lead to the weight vectors:

\[
(F_{RA}, F_{SA}, F_{TA}, F_{UA}) = (0.19, 0.286, 0.238, 0.286)
\]

\[
(F_{RB}, F_{SB}, F_{TB}, F_{UB}) = (0.471, 0.176, 0.294, 0.059)
\]

\[
(F_{RC}, F_{SC}, F_{TC}, F_{UC}) = (0.286, 0.429, 0.179, 0.107)
\]

\[
(F_{RD}, F_{SD}, F_{TD}, F_{UD}) = (0.176, 0.294, 0.53)
\]

\[
(F_{RF}, F_{SF}, F_{TF}, F_{UF}) = (0.4, 0.3, 0.25, 0.05)
\]

Step 3. The production plan is developed based on the above weights as follows:

Production plan in view of bottleneck A: (0R, 60S, 0T, 120U; Profit: 7200)
Production plan in view of bottleneck B: (70R, 50S, 50T, 0U; Profit: 11100)
Production plan in view of bottleneck C: (70R, 60S, 40T, 0U; Profit: 11200)
Production plan in view of bottleneck D: (0R, 0S, 30T, 150U; Profit: 6000)
Production plan in view of bottleneck F: (70R, 60S, 40T, 0U; Profit: 11200)

To illustrate, consider the product mix in view of B. As R is assigned the highest weight by B, it is produced first. After meeting its demand, the next important product, T, is produced. After producing 50 T, the time left in the bottlenecks is only enough to produce 50 S. The profit is calculated by multiplying the produced amount of each product and its contribution margin:

\[
Pr ofit = 70 \times 80 + 50 \times 60 + 50 \times 50 = 11100
\]

Step 4. Weights of the bottlenecks are calculated by normalizing the profit values:

\[
(W_A', W_B', W_C', W_D', W_F') = (0.154, 0.238, 0.24, 0.128, 0.24)
\]

Step 5. The following model is used to determine the weight of each product assigned by the group of bottlenecks (aggregated weight).

\[
\begin{align*}
\min & \quad 0.154|W_R - 0.19| + 0.238|W_R - 0.471| + 0.24|W_R - 0.286| + 0.128|W_R - 0| \\
& + 0.24|W_R - 0.4| + 0.154|W_S - 0.286| + 0.238|W_S - 0.176| + 0.24|W_S - 0.429| \\
& + 0.128|W_S - 0.176| + 0.24|W_S - 0.3| + 0.154|W_T - 0.238| + 0.238|W_T - 0.294| \\
& + 0.24|W_T - 0.179| + 0.128|W_T - 0.294| + 0.24|W_T - 0.25| + 0.24|W_U - 0.286| \\
& + 0.238|W_U - 0.059| + 0.24|W_U - 0.107| + 0.128|W_U - 0.53| + 0.24|W_U - 0.05|
\end{align*}
\]

Subject to:

\[
W_R + W_S + W_T + W_U = 1
\]

\[
0 \leq W_R \leq 0.471
\]
0.176 \leq W_S \leq 0.429

0.179 \leq W_T \leq 0.294

0.05 \leq W_U \leq 0.53

The model is simplified to a linear one:

\[\text{Min } 0.154(W_{RA}^+ + W_{RA}^-) + 0.238(W_{RB}^+ + W_{RB}^-) + 0.24(W_{RC}^+ + W_{RC}^-) + 0.128(W_{RD}^+ + W_{RD}^-) + 0.24(W_{RF}^+ + W_{RF}^-) + 0.154(W_{SA}^+ + W_{SA}^-) + 0.238(W_{SB}^+ + W_{SB}^-) + 0.24(W_{SC}^+ + W_{SC}^-) + 0.128(W_{SD}^+ + W_{SD}^-) + 0.24(W_{SF}^+ + W_{SF}^-) + 0.154(W_{TA}^+ + W_{TA}^-) + 0.238(W_{TB}^+ + W_{TB}^-) + 0.24(W_{TC}^+ + W_{TC}^-) + 0.128(W_{TD}^+ + W_{TD}^-) + 0.24(W_{TF}^+ + W_{TF}^-) + 0.154(W_{UA}^+ + W_{UA}^-) + 0.238(W_{UB}^+ + W_{UB}^-) + 0.24(W_{UC}^+ + W_{UC}^-) + 0.128(W_{UD}^+ + W_{UD}^-) + 0.24(W_{UF}^+ + W_{UF}^-)\]

Subject to:

\[W_{RA}^+ - W_{RA}^- = W_R - 0.19\]
\[W_{RB}^+ - W_{RB}^- = W_R - 0.471\]
\[W_{RC}^+ - W_{RC}^- = W_R - 0.286\]
\[W_{RD}^+ - W_{RD}^- = W_R\]
\[W_{RF}^+ - W_{RF}^- = W_R - 0.4\]
\[W_{SA}^+ - W_{SA}^- = W_S - 0.286\]
\[W_{SB}^+ - W_{SB}^- = W_S - 0.176\]
\[W_{SC}^+ - W_{SC}^- = W_S - 0.429\]
\[W_{SD}^+ - W_{SD}^- = W_S - 0.176\]
\[W_{SF}^+ - W_{SF}^- = W_S - 0.3\]
\[W_{TA}^+ - W_{TA}^- = W_T - 0.238\]
\[W_{TB}^+ - W_{TB}^- = W_T - 0.294\]
\[W_{TC}^+ - W_{TC}^- = W_T - 0.179\]
\[W_{TD}^+ - W_{TD}^- = W_T - 0.294\]
An Efficient Algorithm Based on TOC to Solve Product Mix Problems

\[ W_{TF}^+ - W_{TF}^- = W_F - 0.25 \]
\[ W_{UA}^+ - W_{UA}^- = W_U - 0.286 \]
\[ W_{UB}^+ - W_{UB}^- = W_U - 0.059 \]
\[ W_{UC}^+ - W_{UC}^- = W_U - 0.107 \]
\[ W_{UD}^+ - W_{UD}^- = W_U - 0.53 \]
\[ W_{UF}^+ - W_{UF}^- = W_U - 0.05 \]
\[ W_R + W_S + W_T + W_U = 1 \]
\[ 0 \leq W_R \leq 0.471 \]
\[ 0.176 \leq W_S \leq 0.429 \]
\[ 0.179 \leq W_T \leq 0.294 \]
\[ 0.05 \leq W_U \leq 0.53 \]
\[ W_{ij}^+, W_{ij}^- \geq 0 \]

The optimum solution is:

\((W_R, W_S, W_T, W_U) = (0.343, 0.3, 0.25, 0.107)\)

Step 6. The production plan regarding the aggregated weights is 70R, 60S, 40T and 0U and total profit is 11200 dollars.

Step 7. As the priority of products are R, S, T and U, allowable tradeoffs are shown in Table 4.

<table>
<thead>
<tr>
<th>Decrease from</th>
<th>Increase to</th>
<th>K value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>R</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>R</td>
<td>U</td>
<td>3</td>
</tr>
<tr>
<td>S</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>S</td>
<td>U</td>
<td>3</td>
</tr>
<tr>
<td>T</td>
<td>U</td>
<td>2</td>
</tr>
</tbody>
</table>

Step 8. To illustrate the remaining steps, consider decreasing T and increasing U (the complete calculations are in Table 5). As contribution margins of T and U are 50 and 30 respectively, two units of U should be increased in turn of decreasing one unit of T. So \(k=2\) for this tradeoff.

Step 9. Now the question is that how many times we can decrease one unit of T and increase two units to U. According to the product mix developed in Step 6, the time left in bottlenecks A to F are 0, 1250, 1000, 0, 1550 minutes. Decreasing one unit of T causes that 10, 5, 10, 15 and 5 minutes are added to the time left in bottlenecks and increasing two units to U consumes 10, 30, 20, 10 and 30 minutes. According to bottleneck A, the iterations can be done 40 times:
\[ k_{tUA} - t_{TA} = 2 \times 5 - 10 = 0 \]

\[ H_A = \min(P_T, \left(\frac{D_U - P_U}{k}\right)) = \min(40, \left(\frac{150 - 0}{2}\right)) = 40 \]

Similarly, the following values are calculated:

\[ k_{tUB} - t_{TB} = 2 \times 15 - 5 > 0 \]

\[ H_B = \min(P_T, \left(\frac{D_U - P_U}{k}\right), \left(\frac{L_B}{k_{tUB} - t_{TB}}\right)) = \min(40, \left(\frac{150 - 0}{2}\right), \left(\frac{1250}{2 \times 15 - 5}\right)) = 40 \]

\[ k_{tUC} - t_{TC} = 2 \times 10 - 10 > 0 \]

\[ H_C = \min(P_T, \left(\frac{D_U - P_U}{k}\right), \left(\frac{L_C}{k_{tUC} - t_{TC}}\right)) = \min(40, \left(\frac{150 - 0}{2}\right), \left(\frac{1000}{2 \times 10 - 10}\right)) = 40 \]

\[ k_{tUD} - t_{TD} = 2 \times 5 - 15 < 0 \]

\[ H_D = \min(P_T, \left(\frac{D_U - P_U}{k}\right)) = \min(40, \left(\frac{150 - 0}{2}\right)) = 40 \]

\[ k_{tUF} - t_{TF} = 2 \times 15 - 5 > 0 \]

\[ H_F = \min(P_T, \left(\frac{D_U - P_U}{k}\right), \left(\frac{L_F}{k_{tUF} - t_{TF}}\right)) = \min(40, \left(\frac{150 - 0}{2}\right), \left(\frac{1550}{2 \times 15 - 5}\right)) = 40 \]

\[ H = \min(\{H_A, H_B, H_C, H_D, H_F\}) = 40 \]

\( H \) values for other tradeoffs, that are calculated in a similar way, are zero.

Step 10. Decreasing 40 units from T and increasing 80 units to U will increase the profit value for 400 dollars. As \( H \) values for other tradeoffs are aero, the net profit change according to them are zero. As a result, decreasing T and increasing U is selected and the improved plan is 70R, 60S, 0T and 80U and the profit will be 11600.

Steps 8 to 10 are repeated until no profitable tradeoff is left. The results are summarized in Table 5.

<table>
<thead>
<tr>
<th>Selected Tradeoff</th>
<th>Decrease from</th>
<th>Increase to</th>
<th>K</th>
<th>H</th>
<th>Net change</th>
<th>Product mix</th>
<th>Total profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>U</td>
<td>2</td>
<td>40</td>
<td>400</td>
<td>70R, 60S, 0T, 80U</td>
<td>11600</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>T</td>
<td>2</td>
<td>6</td>
<td>120</td>
<td>64R, 60S, 12T, 80U</td>
<td>11720</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>U</td>
<td>2</td>
<td>8</td>
<td>80</td>
<td>64R, 60S, 4T, 96U</td>
<td>11800</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>T</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>62R, 60S, 8T, 96U</td>
<td>11840</td>
<td></td>
</tr>
</tbody>
</table>

So, the final plan is 62R, 60S, 8T and 96U and the profit is 11840. The result of the proposed algorithm is compared to the traditional, revised and improved algorithms as well as ILP model in Table 6.
Table 6. Comparison of the results

<table>
<thead>
<tr>
<th>Product mix</th>
<th>Traditional Algorithm</th>
<th>Revised Algorithm</th>
<th>Improved Algorithm</th>
<th>Proposed Algorithm</th>
<th>ILP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>70R, 50S, 50T, 0U</td>
<td>64R, 2S, 50T, 120U</td>
<td>70R, 60S, 80T, 0U</td>
<td>62R, 60S, 8T, 96U</td>
<td>51R, 38S, 50T, 100U</td>
<td>51R, 38S, 50T, 100U</td>
</tr>
<tr>
<td>Profit</td>
<td>11100</td>
<td>11340</td>
<td>11600</td>
<td>11840</td>
<td>11860</td>
</tr>
</tbody>
</table>

6. SUMMERY

A number of algorithms have been developed on determining product mix under TOC. Due to the inefficiency of these algorithms, a new algorithm based on a different and new approach is developed. This paper considers product mix problem as a group decision making problem and benefits from using a linear model to aggregate products weights and finds an initial solution. Then tries to improve the solution based on a new logic for decrease and increase process. Using the linear model ensures that the number of alternatives for decrease and increase is decreased significantly. This helps the algorithm to reach the final solution in a short time. Comparison of the proposed algorithm with the traditional, revised and improved ones shows that the product mix resulted by the proposed algorithm is more desirable than all previous algorithms.

REFERENCES

Case Study of Power System Faults in the Petroleum Development Oman (PDO) Network

Wenping Cao*, Nabil Said Al-Kindi
School of Science and Technology
University of Teesside
Middlesbrough, England, TS1 3BA

ABSTRACT

This study is to investigate power system faults in the 33 kV Fahud area under Petroleum Development Oman (PDO) power network. The sequence bus impedance method is used in Matlab environment. The impacts and vulnerability of each bus of interest are investigated so as to find the weakest buses which will need particular attention in power system protection consideration. The main objective is to assess four types of faults as per their levels of fault occurrence and impacts on different buses in the power system so that the behaviour of the power system can be then improved after gathering the correct system information and evaluating results of the faults analysis. Calculation results are confirmed by the practical failures and failure rates.

1. INTRODUCTION

The power system of the Petroleum Development Oman (PDO) is a significant part of electricity supply in Oman, primarily generated from oil and gas. Situated in Fahud and Natih areas, the system consists of 132 kV and 33 kV overhead transmission lines. Each area is supplied from 33 kV switch gears through a transmission line ending to a 33 kV switchyard, then the voltage is stepped down as per the operation requirement.

Fahud sub-system is situated in the northern area under the PDO and is one of the oldest oil fields in Oman. For a while, frequent power faults occur in Fahud area within the PDO power network system, which constantly cause an oil deferment. Among these failure cases, cable termination failures (as shown in Figure 1) occur mostly for 33 kV distribution lines where there is a crossing of graded road. Sometimes these power system failures are due to surge arrestors and 33 kV field isolator damages. When a power failure occurs in this area, an under voltage in the system normally causes a tripping to the circuit breaker, resulting in an escaladed power failure to the location and causing the under voltage relay to activate.

Figure 1: Two cases of cable termination failures of a transmission line within PDO.
(a) Cable insulation failure, (b) Isolator caught flashover.
There are mainly two impacts of the short circuit faults upon the power system, thermal and mechanical effects. Thermal effects on conductors and equipment are created due to heating and excess energy input over time whilst short circuit currents flowing through. Because the short circuit durations are typically short, the heat loss from conductors during the fault may be fairly low. Generally both ac and dc components of the short circuit current contribute to the thermal heating of conductors. Extreme values of the time constant of the dc short circuit current component may be up to 150 ms so that following 500 ms from the instant of the short circuit the dc component is nearly vanished and all its generated heat will have dissipated by around 1s. For the current component, the heat dissipation depends on the ratio of the initial sub-transient rms current to steady state rms current [1].

Power system must be protected against short circuit currents by disconnecting the faulty part of the system via circuit breakers or other breaking apparatus operating in conjunction with protection relays. By doing so, the safe disconnection is guaranteed if the current does not exceed the limit of the circuit breaker. Therefore, the short circuit currents in the network must be analysed beforehand with the precise rating of each appointed system. Short circuit fault calculations should be carried out to ensure the safety of the public as well as the assets. Power system equipment such as circuit breakers are prone to failure with fault duties exceeding their rating. Cables, busbars and transformers are weak thermally or mechanically if subjected to frequent fault currents near their ratings.

2. PDO POWER SYSTEM

The PDO daily requirement for its operation is between 530 MW to around 1400 MW. In 2007, the PDO had about 2400 km of 132 kV overhead transmission lines. There is also an extensive network of 33 kV distribution overhead lines for distributing power to stations, field wells, and utilities. The Fahud sub-system consists of 6 gathering stations and one main pumping station. In addition, there is a booster station which is supplied by the 132 kV transmission line. The power supply in Fahud is transmitted through 33 kV lines across the area. Each feeder is connected to 33 kV switchboards through overhead lines to the located area. Most of the loads in Fahud operate at either 11 kV or 6.6 kV for the large motors and normal 415 volts for the utilities. The Fahud main line pumping station is fed through 33/11 kV while other stations are fed through 33/6.6 kV.

Frequent failures of the power supply in Fahud area have led to costly maintenance and repair. As a result, concerns have been raised about the rationales behind these problems. In economical terms, yearly development is undertaking to improve the oil field production. Compared to other areas of the PDO, there is a significant deferment in Fahud in terms of oil production from the field statistics. This is the main motivation to initiate this project. Though a power fault analysis, short circuit currents at the interested buses in the power system can be computed and compared with the rating of the power system protection equipment such as circuit breakers and relays.

3. SEQUENCE BUS IMPEDANCE METHOD

Analysis of short circuit faults can be conducted using the sequence bus impedance method [2][3]. In an N-bus power system, the system can be modelled by its positive-, negative- and zero-sequence networks where transmission lines and transformers are represented by series reactances, and synchronous machines are represented by constant-voltage sources. If all resistances, shunt admittances, non-rotating loads, and prefault load currents are neglected, the sequence rake equivalents for both symmetrical and unsymmetrical faults at bus n of an N-bus three-phase power system can be obtained. The prefault voltage \(V_i\) is included in the positive-sequence rake equivalent. In this paper, each bus impedance element has an additional subscript 0, 1, or 2 indicating the zero-, positive- and negative-sequence components, respectively.

In order to apply the bus impedance method to the PDO system, a single line diagram for the PDO in Fahud area is produced, as illustrated in Figure 2. In this diagram, the faulty area is simplified to include only 8 buses which have severe fault occurrences in the past and which are of particular interest. Other less significant buses are not numbered in this study so as to reduce the computation effort. This study begins with gathering sufficient data for the existing network which include the buses, bus sequence impedances, and the ratings of transformers and switchgears. The sequence bus impedance method is used in MatLab to evaluate the vulnerability of each bus and the impacts of each faulty bus on any other buses.

Having obtained the strength and weakness of each individual bus, the severity of the power fault strikes can be estimated and ranked in order of significance. The faults considered in this paper consist of balanced 3-phase, single-phase and double-phase fault types. After reviewing the current power system design, some improvements can be
made to protect the weakest buses first. As a result, the overall fault occurrences can be minimised and the system stability be improved.

A simple flow diagram for short circuit calculations is shown in Figure 3. First, all input data are entered and then the sequence and the admittances of each sequence network are identified and labelled in the m file, flowed by converting the admittance matrices into the impedance matrices.
From the theoretical analysis [2], the fault sequence currents for all types of fault at bus $n$ can be computed, as summarised in Table 1. The sequence components of the phase voltages at any bus $k$ during a fault at bus $n$ are found by [3]:

$$ \begin{bmatrix}
V_{k-0} \\
V_{k-1} \\
V_{k-2}
\end{bmatrix} = \begin{bmatrix}
0 & Z_{kn-0} & 0 & 0 \\
V_f & 0 & Z_{kn-1} & 0 \\
0 & 0 & 0 & Z_{kn-2}
\end{bmatrix} \begin{bmatrix}
I_{n-0} \\
I_{n-1} \\
I_{n-2}
\end{bmatrix} = \begin{bmatrix}
-Z_{kn-0}I_{n-0} \\
-V_{kn-1}I_{n-1} \\
-Z_{kn-2}I_{n-2}
\end{bmatrix} $$

(5)

Table 1: Fault sequence current calculations
4. RESULTS AND DISCUSSIONS

After obtaining the sufficient bus information, a set of sequence admittance matrices are formed with a matrix size of 8×8. Eight buses are assigned for short circuit fault assessments with a base power of 50 MVA. Using Equations 1-4, the sequence fault currents for the fault bus \( N \) can be computed. As shown in Figure 3, by inverting the three sequence admittance matrices, the corresponding bus impedance matrices can be given (thus the name bus impedance method). Using Equation 5, the bus voltages for all buses can be obtained and so are the phase currents, active and reactive powers. These calculations are complex and therefore conducted in MatLab environment.

4.1 Balanced three phase fault

First, a balanced three phase faults is analysed. The impact of a symmetrical three-phase fault at each bus \( n \) on the eight buses \( k \) are calculated using the sequence bus impedance method. The fault currents and voltages are based on a bolted three phase fault at bus 1 to bus 8. Fault current levels are shown in Figure 4 with the fault bus being 1 to 8 respectively. From Figure 4, it can be seen that the fault current would be significant if the balanced fault occurs at bus 5, 6, 8, 1 in descending order. Clearly this fault represents the most severe type of the faults in the system. However, it does not occur frequently in the PDO system.

![Figure 4: Balanced 3 phase fault current levels.](image)

4.2 Single phase to ground fault
The most frequent type in service is still the single phase to ground fault. Therefore a single phase (phase A) to ground fault is applied and the calculation results from the bus impedance method are given in Table 2.

Table 2: Bus phase voltages during a single-phase-ground fault (pu magnitude/angle)

<table>
<thead>
<tr>
<th>Fault bus</th>
<th>Bus evaluated</th>
<th>Phase A</th>
<th>Phase B</th>
<th>Phase C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1.2011/0</td>
<td>1.2011/-120</td>
<td>1.2011/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.2896/0</td>
<td>1.2896/-120</td>
<td>1.2896/120</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.0303/0</td>
<td>1.0303/-120</td>
<td>1.0303/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.0573/0</td>
<td>1.0573/-120</td>
<td>1.0573/120</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.0305/0</td>
<td>1.0305/-120</td>
<td>1.0305/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.0265/0</td>
<td>1.0265/-120</td>
<td>1.0265/120</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.1272/0</td>
<td>1.1272/-120</td>
<td>1.1272/120</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.0290/0</td>
<td>1.0290/-120</td>
<td>1.0290/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3418/0</td>
<td>1.3418/-120</td>
<td>1.3418/120</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.1158/0</td>
<td>1.1158/-120</td>
<td>1.1158/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.1214/0</td>
<td>1.1214/-120</td>
<td>1.1214/120</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.0965/0</td>
<td>1.0965/-120</td>
<td>1.0965/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.1487/0</td>
<td>1.1487/-120</td>
<td>1.1487/120</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.1987/0</td>
<td>1.1987/-120</td>
<td>1.1987/120</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.0342/0</td>
<td>1.0342/-120</td>
<td>1.0342/120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.2798/0</td>
<td>1.2798/-120</td>
<td>1.2798/120</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.1051/0</td>
<td>1.1051/-120</td>
<td>1.1051/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.1187/0</td>
<td>1.1187/-120</td>
<td>1.1187/120</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.0893/0</td>
<td>1.0893/-120</td>
<td>1.0893/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.1308/0</td>
<td>1.1308/-120</td>
<td>1.1308/120</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.2054/0</td>
<td>1.2054/-120</td>
<td>1.2054/120</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.0110/0</td>
<td>1.0110/-120</td>
<td>1.0110/120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.2909/0</td>
<td>1.2909/-120</td>
<td>1.2909/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3225/0</td>
<td>1.3225/-120</td>
<td>1.3225/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.2276/0</td>
<td>1.2276/-120</td>
<td>1.2276/120</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.2182/0</td>
<td>1.2182/-120</td>
<td>1.2182/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.3866/0</td>
<td>1.3866/-120</td>
<td>1.3866/120</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.2893/0</td>
<td>1.2893/-120</td>
<td>1.2893/120</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.0173/0</td>
<td>1.0173/-120</td>
<td>1.0173/120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.2547/0</td>
<td>1.2547/-120</td>
<td>1.2547/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3041/0</td>
<td>1.3041/-120</td>
<td>1.3041/120</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.1900/0</td>
<td>1.1900/-120</td>
<td>1.1900/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.1518/0</td>
<td>1.1518/-120</td>
<td>1.1518/120</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.2599/0</td>
<td>1.2599/-120</td>
<td>1.2599/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.2296/0</td>
<td>1.2296/-120</td>
<td>1.2296/120</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.0122/0</td>
<td>1.0122/-120</td>
<td>1.0122/120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.2682/0</td>
<td>1.2682/-120</td>
<td>1.2682/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3032/0</td>
<td>1.3032/-120</td>
<td>1.3032/120</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.2413/0</td>
<td>1.2413/-120</td>
<td>1.2413/120</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.2010/0</td>
<td>1.2010/-120</td>
<td>1.2010/120</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.3351/0</td>
<td>1.3351/-120</td>
<td>1.3351/120</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.2603/0</td>
<td>1.2603/-120</td>
<td>1.2603/120</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.0088/0</td>
<td>1.0088/-120</td>
<td>1.0088/120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.3433/0</td>
<td>1.3433/-120</td>
<td>1.3433/120</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3689/0</td>
<td>1.3689/-120</td>
<td>1.3689/120</td>
</tr>
</tbody>
</table>
From this table, it can be observed that, among all these 8 buses of interest, buses 2, 3, 4, 7 and 8 are likely to be the weakest buses in reducing significance for this type of fault. These buses are found to be problematic in most of the distribution lines. Having checked with the past failure rates, there is a positive correlation between actual failures and these buses identified. In practice, the failures are mostly due to power cable termination failures, lighting arrestors and some types of 33 kV insulators.

4.3 Double phase fault

Phase to phase to ground faults are shown in Figure 5. They are in similar order to the balanced three phase fault.

![Diagram of Fault Buses](image1)

**Figure 5**: Phase to phase fault current magnitudes.

4.4 Double phase to ground fault

Phase to phase to ground faults are shown in Figure 6. They are in similar order to the balanced three phase fault.

![Diagram of Fault Buses](image2)

**Figure 6**: Phase to phase to ground fault current magnitudes.

4.5 Fault power contribution
In terms of fault power contribution, each fault results in different fault power levels. Shown in Figure 7 is a summary of all types of fault with all bus being faulty in turn.

From this figure, the significance of power contributions actually changes with the fault type. If compared to the previous results for fault currents, the buses are not in the similar order of significance. As a result, it may be necessary to consider the different fault current/power levels in protection coordination, taking account of the past history of fault types in the Fahud area.

5. SUMMARY

A case study of the PDO power system has been presented. Short circuit faults are studied including the most severe type (balance three-phase) and the most frequent type (the single to phase to ground). High levels of fault currents and powers can result in thermal and mechanical stresses on the power components and thus cause damage to operational equipment or even to the personnel in the environment.

Using the sequence bus impedance method, the Fahud sub-system is simplified to be an 8-bus single-diagram system for power fault analysis. By computing each individual bus’s impacts on other buses and its vulnerability with reference to a particular type of faults, the significance of the bus can be assessed in terms of its strength and weakness. This information can be used for considering enhancement of existing power system protection. The fault current results on the 8 buses have been confirmed by the actual causes of power failures.

This power fault analysis will lead to a revision of the PDO network and future development plan, based on the past statistics of power system fault and frequency. A further investigation is required using a power flow programme to investigate the Fahud 33 kV power system stability caused by these faults.

REFERENCES

A new strategy for resource-constrained allocation in PERT networks under the condition of renewable resource

Siamak Baradaran a, S. M. T. Fatemi Ghomi b,*, and S.S. Hashemin c

a Institute for Trade Studies and Research
240 North Kargar Avenue, Tehran, Iran

b Department of Industrial Engineering
Amirkabir University of Technology
424 Hafez Avenue, Tehran, Iran

c Department of Industrial Engineering
Islamic Azad University
Arda Branch, Ardebil, Iran

ABSTRACT

The complexity of NP-hard optimization problems caused using heuristic solution procedures. The resource-constrained allocation problem (RCAP) in PERT type networks belongs to this class. Therefore, using these procedures is indispensable. Few algorithms are presented in this area in stochastic networks. Most of these algorithms are based on developing and carrying out several strategies to prioritize the activities which must be operated at decision points. In this paper, a new strategy is suggested via traditional strategies in resource management. The three main kinds of strategies were used in literature, have been composed; activity oriented, resource oriented and network oriented. This composed strategy has been used for considered problem, which is the scheduling of project activities subject to finish-start precedence constraints with zero time lag and constant renewable resource constraints in order to minimize the project duration in PERT networks. The new composite strategy performed well, compared to well-known traditional strategies.

1. INTRODUCTION

We will consider a RCAP as follows. A single project consists of a set of activities which have to be processed. The activities are interrelated by two kinds of constraints. First, precedence constraints force each activity not to be started before all its immediate predecessor activities comprised in the set have been finished. Second, performing the activities requires renewable resources with deterministic limited capacities at any decision point, which must be operated prioritize based on strategies. The resource constraints refer to limited renewable resources such as manpower, material and machines which are necessary for carrying out the project activities. Where resources are renewable, activity durations are resource driven [1]. The activity durations are independent continuous random variables; preemption is not allowed. The objective of the problem is the feasible sequence for project activities such that makespan of project be minimized.

Blazewicz et al. [2] have shown that this problem is a generalization of the classical job shop scheduling problem and belongs to the class of NP-hard optimization problems. Therefore, heuristic solution procedures are indispensable when solving large problem instances as they usually appear in practical cases. Since 1963 when Kelley [3] introduced a schedule generation scheme, a large number of different heuristic algorithms have been suggested in the literature [4].

1.1. HEURISTIC IN SCHEDULING

The limited resource project scheduling problem falls into a category of mathematical problems known as combinatorial problems. This is because, for any given problem, a very large number of possible combinations of activity start times exist, with each combination representing a different project schedule. The number of

* Corresponding author: Tel.: (+9821) 66413034; Fax: (+9821) 66413025; E-mail: fatemi@aut.ac.ir
combinations is extremely large, and increases rapidly with an increase in the number of activities. In fact, the number is typically so large for realistic-sized problems as to prohibit enumeration of all alternatives, even with the aid of a computer. Analytical methods such as mathematical programming have not proven very successful on these combinatorial problems. Instead, various heuristic-based procedures have been developed [5]. Heuristics may not always produce the best solution in every case, but their usefulness in finding good solutions with a minimum of effort is well-known, based on experience and research studies. The heuristic procedures broadly fall into two categories, namely constructive heuristics and improvement heuristics [6]. Constructive heuristics consist of two major components, namely the scheduling scheme and the priority rule. The scheduling scheme determines the way in which a feasible schedule is constructed by assigning starting times to the different activities. The priority rule determines the activity that is selected next during the heuristic search process.

Many researchers have investigated the resource allocation problem; Brucker et al. [7] and Ozdamar and Ulusoy [8] present extensive overviews of resource allocation research. Strategies and priority rules can be classified in three kinds of strategies are used in literature [9]. The first group is the activity oriented (base) priority rules which consider information related to the activity; the second is the resource oriented (base) priority rules which consider the resource use of the different activities. The third is the network oriented (base) priority rules which only consider the information contained in the network itself, not to the rest of the project.

The three main kinds of strategies were used in literature, have been composed and applied this composed strategy for considered problem. The new composite strategy compared to well-known traditional strategies. According to the classification scheme for project scheduling from Herroelen et al. [10], a regular performance measure is used, in particular the makespan. That is to say, the objective is to minimize the project makespan. Instead of analyzing the merely makespan values, the percentage increase in the makespan values with respect to the best possible solution project duration has been analyzed. Thus, this is a measurement of the efficiency, given that the performance of the new combination has been compared with the best possible solution.

The paper is organized as follows. Section 2 deals with the problem description. Section 3 gives the development of new strategy for RCAP in PERT networks “RAP”. Experimental design is treated in Section 4. Section 5 is devoted to the discussion and the paper has been ended with the conclusions in Section 6.

2. THE PROBLEM FORMULATION

The well-known resource-constrained allocation problem (RCAP) can be stated as follows:

Project network G(N, A) is shown as PERT type network with an activity-on-arc (AOA) format, the set of arcs A are used to represent the n activities (numbered from 1 to n, i.e.) and a set of nodes N represent the precedence relations between activities, which are assumed to be finish-start relations with a minimal time-lag of zero. Activity durations are independent continuous random with given distribution function which requires only renewable resource. The amount of limited available resource is deterministic and allocated is performed discretely. The activities are assumed to be nonpreemptable. A regular performance measure is used, in particular the makespan. That is to say, the objective is to minimize the project makespan.

According to above, the considered model is defined as follows:

\[
\begin{align*}
G(N,A) & \quad \text{stochastic network (graph) of PERT type with A activities and N as the set of precedence relationships} \\
(i, j) \in A & \quad \text{the project’s activity } (i, j) \\
F_i & \quad \text{the finish time of activity } (i, j) \\
R_k\left(t / F_i\right) & \quad \text{maximal value of the } k \text{-th resource profile at point of } t \text{ on condition that activities } (i, j) \in A \text{ are finished at point of } t \\
T(G) & \quad \text{random variable of project completion time} \\
R_k & \quad \text{total available resource of type } k \text{ at the project management disposal (pre-given and fixed throughout the planning horizon)} \\
K & \quad \text{number of different resources} \\
SAS & \quad \text{scheduled activity set}
\end{align*}
\]
According to the notations, the considered model is defined as follows:

\[
\text{Min}_{ij} f( E(T(G)/ F_{ij}) ) \\
\text{s.t.: } R_k^e(t/F_{ij}) \leq R_k, \\
\quad 1 \leq k \leq K \quad \forall t \geq 0.
\] (1)

The problem is to determine the values of \( F_{ij} \) in order to minimize makespan of project. The set of constraints in this model shows that at each decision point \( t \), the maximum required amount of resource \( k \) is equal or less than its available amount. Model (1) is a stochastic optimization model which cannot be solved in the general case due to its intractability; hence, only a heuristic solution can be applied.

3. THE DEVELOPMENT OF NEW STRATEGY FOR RCAP IN PERT NETWORKS

It seems the strategies are needed for activity selection at decision points. Each priority rule has a characteristics and predictable behavior. Obviously when the priority rules are combined together, the behavior would be changed and new behavior depends on what kind of combination has been made. The strategy can be applied with simple or combination form.

In simple form priority rules may be used in the single or sequential form. In sequential form some priority rules are used as sequential at several steps. For combination of rules, two kinds of operators can be employed. They are summation operator and multiplication operator. When kinds of priority rules are combined with summation operator, each priority rule can be used in simple or weighted form. For combination with multiplication operator, each priority rule can be used in normal or reverse form.

The main purpose of this paper is to use synergy of rules combination. There are kinds of composite rules that their efficiency must be evaluated to select the best possible solution. As mentioned previously, according to the literature [9] the priority rules can be classified in three major groups, activity oriented, resource oriented and network oriented. It can be considered Resource Criticality Index (\( RCI \)), Activity Time Density (\( ATD \)) and Path Index (\( PI \)), respectively as representative of the mentioned groups. Resource Criticality Index (\( RCI \)) is denoted by \( \text{Max}_{ij \in SAS} (R_{ij} \text{activity resource requirement}) / \text{Total resource availabilities} \) and calculated by the following relation:

\[
RCI = \frac{\text{Activity resource requirement}}{\text{Total resource availabilities}}
\] (2)

Activity Time Density (\( ATD \)) is denoted by \( \text{Max}_{ij \in SAS} (ATD_{ij}) \) and calculated by the following relation:

\[
\text{ATD} = \frac{\text{Activity duration}}{\text{Activity duration} + \text{Activity slack}}
\] (3)

It is important to say in this paper stochastic networks are considered with activity continuous time. Therefore generalization of dynamic index has been needed like Activity Time Density (\( ATD \)). To calculate \( ATD \), it is required to calculate the activity duration and the activity slack. However the durations and slacks of activities in PERT type networks are random variables. Their average has been applied estimated by simulation. This rule is noted by \( \text{Max} \ ATD \). Indeed the activities duration times have been generated randomly and simulated the network \( M \) times and computed \( ATD \) with following relation:

\[
\overline{ATD}_q = (\sum_{ij \in A} ATD^q_{ij}) / M \quad (i, j) \in A \quad \text{and} \quad q = 1,2,...,M
\] (4)

Activity Path Index (\( PI \)) is noted by \( \text{Max}_{ij \in SAS} (PI_{ij}) \) and is the number of paths which pass from activity.

As mentioned previously, activities in PERT type networks is random variable. Their average has been applied estimated by simulation. Based on above conclusion the network makespan has been computed by simulation. The
mean of makespan has been computed for all of problems through applying each rule in each type of networks denoted by $\overline{MKSP}$.

$$\overline{MKSP}_q = \left( \sum_{i,j}^n \text{Makespan}^q_{ij} \right) / M \quad (i, j) \in A \quad \text{and} \quad q = 1, 2, \ldots, M$$ (5)

The new strategy is constructed with the combination of three rules; Resource Criticality Index (RCI), Activity Time Density (ATD) and Path Index (PI) that is called “RAP”. The performance of the new strategy is investigated by designing experiments and evaluation of each combination of schemes.

4. EXPERIMENTAL DESIGN

The RAP has been analyzed with respect to the solutions obtained with the best heuristic method in literature. The evaluation method is based on the design of experiments.

4.1. FOUR TYPES OF NETWORK STRUCTURE

The experience gained in resource-constrained allocation with heuristic and optimization procedures has clearly shown that the effectiveness of both types of procedures is strongly dependent upon the characteristics of the particular problem being solved. It has been attempted to have different configurations for network structure considered in the evaluating the alternatives. Having attention to reference Kurtulus and Davis [11], four different types of structures have been considered for networks of this paper. Figure 1 illustrates these structures.

In type A the number of parallel activities is more in the initial portion of network and less in the final portion (Figure 1.a) while in type B it is less in the initial portion of network and more in the final portion (Figure 1.b). For type C the numbers of parallel and independent paths are high in the network (Figure 1.c) and type D the number of parallel activities, in the initial or final portion of network does not make so much difference. The paths are not independent of each other and have common activities (Figure 1.d).

**Figure 1: Four types of network structure**

4.2. EXPERIMENTAL DESIGN

Each activity has the same duration and the same renewable resource requirements in different network types. A designed experiment was used to determine the effect of final three forms of how to apply priority rules in network types. In all cases, four replicates of each treatment combination were made. Therefore (4 different types of A, B, C and D) × [(3 single form) + (6 sequential form) + (7 summation operator) + (7 multiplication operator)] = 92 problems were constructed and solved. The activities durations of the test networks have been randomly generated with Uniform distribution function. To solve all 92 problems, they were coded in Visual Basic 6.0.

The combination schemes have been evaluated for the two specific cases. The performance of the new composite strategy was investigated by comparing solutions of the different combination schemes in two specific cases. Indeed
the results are compared with the other solutions especially with well-known traditional strategy, the Time Criticality Index \( (TCI) \) which is denoted by \( \text{MaxTCI} \) and calculated by the following relation:

\[
TCI = \frac{\text{No. of criticality status for each activity}}{M} \times 100
\]  

where \( M \) is the number of simulation runs.

The number of criticality status has been applied for each activity in simulation runs and calculated the percentage of criticality. Indeed the activities duration times have been generated randomly and simulated the network \( M \) times and computed \( TCI \). In the first case, combination schemes of \( \text{RAP} \) efficiency is measured against the best possible solution (BPS) that is the minimum of \( \text{MKSP} \) for each network type. As mentioned previously, in this case the best possible solution is selected and compared with the other solutions especially with the Time Criticality Index \( (TCI) \) solution. Therefore, by resolving the resource conflicts with the \( \text{MaxTCI} \) in comparison to the best possible solution the validation of our index is approved, which is expressed in percentage denoted by “Index of Best Possible Solution” (IBPS) as follows:

\[
\text{IBSP} = \frac{\text{MKSP} - \text{Best Possible Solution}}{\text{Best Possible Solution}} \times 100
\]

Also the efficiency of the single project environment in larger projects can be measured with regard to critical path \[12\]. Therefore in the second case efficiency of new combined priority rule \( \text{RAP} \) is measured against mean of “Initial Critical Path” which is mean of makespan by \( M \) times network simulation with non-constraint resources. Therefore it is expressed in percentage the difference of makespan in two situations and denoted by “Index of Critical Path” (ICP) as follows:

\[
\text{ICP} = \frac{\text{MKSP} - \text{Initial Critical Path}}{\text{Initial Critical Path}} \times 100
\]

Additionally in this case the initial critical path solution has been compared with other solutions and especially with the \( \text{MaxTCI} \) solution in single form of priority rules. Thus, by resolving the resource conflicts with the \( \text{MaxTCI} \) in comparison to other solutions the validation of our selected index is approved. The two mentioned cases allow us to correctly evaluate efficiency of \( \text{RAP} \).

5. DISCUSSION

The \( \text{RAP} \) has been analyzed with respect to the solutions obtained with one of the best heuristic method in literature. As mentioned, four different types of A, B, C and D have been considered for experiments. The first results are shown in Figure. 2 which illustrates the mean of makespan (\( \text{MKSP} \)) values to compare efficiency of \( \text{RAP} \) forms for all selected networks. The experiments and analysis are divided into two cases to compare efficiency of the schedules when the project resources have limited availabilities. The first case is formed by the best possible solution and the second case is formed by initial critical path. Table 1 contains the values of networks solutions to evaluate the \( \text{RAP} \) in resource-constrained project scheduling. In the following section the results have been explained and analyzed \( \text{RAP} \) efficiency by behavior networks analysis and behavior indices analysis in details.

As mentioned previously, the main purpose of this paper is to use synergy of rules combination. Special attention should be made of the interactions between networks types with \( \text{MaxTCI} \), which showed a new light on the previous results. Although according to the literature \( \text{MaxTCI} \) is well-known and one of the best priority rules, in single form its efficiency is not higher than combined rules. As Figure.3 demonstrates, it has the rank of twentieth for network type A, third for network type B, fourteenth for network type C and twenty third for network type D. In none of four network types it gained the first rank.

Figure. 3 illustrates and compares efficiency of \( \text{MaxTCI} \) against the other considered rules in four networks types. This figure shows in type A, 83.33% solutions are better than \( \text{MaxTCI} \), and in type B this percent is 8.33%.
While there are 54.16% and 91.66% for types C and D. These percents refer to high efficiency of MaxTCI in type B and low efficiency in type D. All of these results illustrate the synergy of rules combination. Indeed results confirm that combinations of heuristics can be used to find better solutions to resource allocation problems than single heuristics. As the results indicate, for type A, 78.26% indices have been found the best possible solution.

![Network makespan in different indices](image)

**Figure 2:** Behaviors of mean of makespan (MKSP) in four networks types

These are 8.70% in single form, 26.09% in sequential form, 30.43% in summation form and 13.04% in multiplication form with project makespan 44.89. In type B there is 8.70%. They are 4.35% in summation form and 4.35% in multiplication form with project makespan 47. For type C there are 4.35% in single form, 13.04% in sequential form, 17.39% in summation form and 8.70% for multiplication form with project makespan 48.89. In type D there are 4.35% for single form, 8.70% in sequential form and 30.43% for summation form and 4.35% for multiplication form with project makespan 44.85.

![Difference among MaxTCI and other indices in four networks types](image)

**Figure 3:** Difference among MaxTCI and other indices in four networks types

Regarding the interaction RAP by four networks, as mentioned, the analysis is performed for two cases for comparison of the schedules when the project resources have limited availabilities. The first case is formed by the best possible solution, which presents a high performance and the second case is formed by critical path. Table 1 shows the percent differences among the values (MKSP) with the best possible solution (minimum of MKSP) and initial critical path (ICP). Table 1 gives the differences clearly. The interpretation is that comparing the MKSP with index of critical path (ICP), in type A, there are 49.58% difference between the best possible solution and initial critical path, and 92.77% difference between the worst solution and initial critical path. It means that there are range values around 43.19 differences. In type B there is a range percent difference between the best possible solution (94.54%) and worst solution (110.64 %) about 16.10 % difference. For type C there is the range between 69.87% and 87.63% that indicates around 17.76 % difference. Finally there is 25.89 % difference between the best possible solution (45.15%) and the worst solution (71.04%) in type D. With respect to the above
results it appears that the differences between the network types of B and C are similar and network type D is almost similar to both of them; while results of network type A is different with others. Regarding the best possible solution (IBPS) and index of critical path (ICP) percents, it can be said that the network structure plays an important role in efficiency of each index.

Table 1: The results of IBPS and ICP

<table>
<thead>
<tr>
<th>No.</th>
<th>Indices</th>
<th>Type A</th>
<th></th>
<th>Type B</th>
<th></th>
<th>Type C</th>
<th></th>
<th>Type D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IBPS%</td>
<td>ICP%</td>
<td>IBPS%</td>
<td>ICP%</td>
<td>IBPS%</td>
<td>ICP%</td>
<td>IBPS%</td>
<td>ICP%</td>
</tr>
<tr>
<td>1</td>
<td>TCI</td>
<td>11.14</td>
<td>66.24</td>
<td>1.89</td>
<td>98.22</td>
<td>2.05</td>
<td>73.35</td>
<td>11.48</td>
<td>61.81</td>
</tr>
<tr>
<td>2</td>
<td>RCI</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>3</td>
<td>ATD</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>10.45</td>
<td>87.63</td>
<td>4.28</td>
<td>51.36</td>
</tr>
<tr>
<td>4</td>
<td>P1</td>
<td>28.87</td>
<td>92.77</td>
<td>3.94</td>
<td>102.19</td>
<td>1.96</td>
<td>73.21</td>
<td>11.15</td>
<td>61.33</td>
</tr>
<tr>
<td>5</td>
<td>RAP</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>10.45</td>
<td>87.63</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>6</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>10.45</td>
<td>87.63</td>
<td>10.97</td>
<td>61.07</td>
</tr>
<tr>
<td>7</td>
<td>RAP-</td>
<td>28.87</td>
<td>92.77</td>
<td>0.00</td>
<td>94.54</td>
<td>0.00</td>
<td>69.87</td>
<td>17.84</td>
<td>71.04</td>
</tr>
<tr>
<td>8</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>10.45</td>
<td>87.63</td>
<td>0.33</td>
<td>45.63</td>
</tr>
<tr>
<td>9</td>
<td>RAP-</td>
<td>11.05</td>
<td>66.11</td>
<td>4.26</td>
<td>102.81</td>
<td>1.96</td>
<td>73.21</td>
<td>11.15</td>
<td>61.33</td>
</tr>
<tr>
<td>10</td>
<td>RAP+</td>
<td>11.14</td>
<td>66.24</td>
<td>6.15</td>
<td>106.50</td>
<td>10.45</td>
<td>87.63</td>
<td>4.79</td>
<td>52.10</td>
</tr>
<tr>
<td>11</td>
<td>RAP-</td>
<td>11.05</td>
<td>66.11</td>
<td>6.38</td>
<td>106.95</td>
<td>0.00</td>
<td>69.87</td>
<td>0.33</td>
<td>45.63</td>
</tr>
<tr>
<td>12</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>8.28</td>
<td>110.64</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>13</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>10.45</td>
<td>87.63</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>14</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>10.45</td>
<td>87.63</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>15</td>
<td>RAP-</td>
<td>0.00</td>
<td>49.58</td>
<td>0.00</td>
<td>94.54</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>16</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>0.22</td>
<td>70.26</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>17</td>
<td>RAP-</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>18</td>
<td>RAP+</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>19</td>
<td>RAP-</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>0.00</td>
<td>69.87</td>
<td>0.33</td>
<td>45.63</td>
</tr>
<tr>
<td>20</td>
<td>RPA</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>21</td>
<td>ARP</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>10.45</td>
<td>87.63</td>
<td>4.28</td>
<td>51.36</td>
</tr>
<tr>
<td>22</td>
<td>APR</td>
<td>0.00</td>
<td>49.58</td>
<td>6.38</td>
<td>106.95</td>
<td>10.45</td>
<td>87.63</td>
<td>4.28</td>
<td>51.36</td>
</tr>
<tr>
<td>23</td>
<td>PRA</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>0.00</td>
<td>69.87</td>
<td>0.00</td>
<td>45.15</td>
</tr>
<tr>
<td>24</td>
<td>PAR</td>
<td>0.00</td>
<td>49.58</td>
<td>6.06</td>
<td>106.33</td>
<td>10.45</td>
<td>87.63</td>
<td>4.28</td>
<td>51.36</td>
</tr>
</tbody>
</table>

The performance of $RAP^{+/-}$ appears in Table 1. The results of $RAP^{+/-}$ are also surprising. As Figure. 4 clearly displays, this index is the best possible solution for all of the four different network types of A, B, C and D considered for design of experiments. Indeed it is common and distinct solution for all of four networks types. $RAP^{+/-}$ Presents the most efficient solution with reference to resource constrained project scheduling. However because of problem complexity, its quality of solutions compared with the optimal solutions is unknown.

Figure 4: Comparing the RAP$^{+/-}$ with other indices in four networks types
6. CONCLUSIONS

The main objective of this paper was to develop a new combined priority rule as a computationally simple method to find good solutions to resource constraint allocation problems (RCAP). The new combined priority rule, developed for stochastic problems, was adapted to the resource allocation problem with activity continuous distribution function and demonstrated on an established test set of resource allocation problems. The new combined priority rule performed well compared to other well-known heuristics like MaxTCI, that according to the literature it is one of the best well-known priority rules. In investigation of 92 problems, the experiments show that, RAP was able to find the best possible solution for all of four types of networks.

Our evaluation has demonstrated that the area in which there are higher differences regarding capabilities and efficiency of the indices for resource-constrained project scheduling, a new combined priority rule (RAP) could be constructed with combination of three rules; Resource Criticality Index (RCI), Activity Time Density (ATD) and Path Index (PI), that its efficiency is superior than each rule. The experiments confirmed the synergy of rules combination and illustrated that combinations of heuristics can be used to find better solutions to resource allocation problems than single heuristics. Some indices only allow the consideration of simple networks, while others also include complex networks. Finally, regarding the criteria for resource allocation to different activities in the resolution of the mentioned problems RAP is more appropriate than other indices in this sense.

To correctly evaluate the efficiency of the schedules obtained with the different indices two designs of experiments have been created, one in the scheduling of constrained resources in single projects with the best possible solution and the other with index of critical path (ICP). In both designs, instances have been generated in which all the relevant project characteristics have been experimented. The results point out the significant differences in performance of the different indices. Thus, RAP was the most efficient because it offers the schedules with the shortest duration, even better than the other indices, such as single form, multiplication form and summation form. It is also remarkably to note that the most efficient indices have a better performance than the classical methods such as MaxTCI.

REFERENCES

FUZZY ELECTRE: A Multi-Attribute Method In Fuzzy Environment

S.M.T. Fatemi Ghomi*, Mehdi Iranpoor, R. Ghasemy Yaghin

Department of Industrial Engineering, Amirkabir University of Technology, Hafez Ave. No. 424, 15916-34311 Tehran, Iran

ABSTRACT

Decision makers often deal with situations, which include incomplete information as vagueness. On the other hand, most of the real-world problems have more than one attribute to be evaluated. Electre is one of the multi-attribute decision making (MADM) methods, which concentrates on the dominance relations among the alternatives. In this paper, we develop Fuzzy Electre method as a multi-attribute decision making approach in fuzzy environment. We suppose that both the rating of options and the weights of criteria are trapezoid fuzzy numbers.

1. INTRODUCTION

In recent years, multi-criteria decision making (MCDM) problems have been widely discussed by researchers.

In real-world situations, because of incomplete or non-obtainable information -for example human judgments including preferences are often vague and he cannot estimate his preference with exact numbers- therefore they usually are fuzzy/imprecise or interval numbers. Fuzzy set theory was first used to solve decision-making problems by Bellman and Zadeh [1]. Laarhoven and Pedrycz extended Saaty's AHP to deal with the imprecision and subjectiveness in the pairwise comparison process [2]. Jahanshahloo et al. extended the concept of TOPSIS to develop a methodology for solving multi-criteria decision-making problems with interval data [3]. Chen further extended the concept of TOPSIS to develop a methodology for solving multi-person multi-criteria decision-making problems in fuzzy environment [4]. Considering the fuzziness in the decision data and group decision-making process, Chen used linguistic variables to assess the weights of all criteria and the ratings of each alternative with respect to each criterion. Aouam et al. introduced a new methodology that combines both the concept of fuzzy outranking and fuzzy attributes to provide a more flexible way for comparing alternatives [5]. Lee built a group decision-making structure model of risk in software development and proposed two algorithms to tackle the rate of aggregative risk in a fuzzy environment by fuzzy sets theory during any phase of the life cycle [6]. Hsu and Chen proposed a method for aggregating individual fuzzy opinions into a group fuzzy consensus opinion [7]. They presented a procedure for aggregating the expert opinions. Herrera et al. presented a consensus model in group decision-making under linguistic assessments [8]. It is based on the use of linguistic preferences to provide individuals’ opinions, and on the use of fuzzy majority of consensus, represented by means of a linguistic quantifier. Li investigated the problems of decision-making with multiple judge, multiple criteria in a fuzzy environment, where the performance of alternatives and the importance of criteria are imprecisely defined and represented by fuzzy sets [9]. A fuzzy model associated with the solution algorithm is proposed on the basis of an $\alpha$-level weighted, fuzzy preference relation.

Electre is one of the other MADM methods, which concentrates the analysis on the dominance relations among the alternatives. That is, this method is based on the study of outranking relations, exploiting notions of concordance. These outranking relations are built in such a way that it is possible to compare alternatives. The information required by Electre consists of information among the criteria and information within each criterion. It is a comprehensive evaluation approach in that it also tries to rank a number of alternatives each one of which is described in terms of a number of criteria. The main idea is the proper utilization of what is called “outranking

* - Corresponding author: Tel.: (009821)66413034 Fax: (009821)66413025; E-mail: fatemi@aut.ac.ir
relations”. Electre is more than just a solution method; it is a philosophy of decision aid. This philosophy is discussed at length by Roy [10]. Electre has evolved through a number of versions (I through IV); all are based on the same fundamental concepts but are operationally somewhat different.

Some of the situations in which Electre methods are more suitable than aggregation methods, are listed below [11].

1. The decision-maker (DM) wants to include in the model at least three criteria. However, aggregation procedures are more adapted in situations when decision models include more than five criteria (up to twelve or thirteen).

2. A strong heterogeneity related with the nature of evaluations exists among criteria (e.g., duration, noise, distance, security, cultural sites, monuments, ...). This makes it difficult to aggregate all the criteria in a unique and common scale.

3. Compensation of the loss on a given criterion by a gain on another one may not be acceptable for the DM. Therefore, such situations require the use of noncompensatory aggregation procedures (e.g. Electre methods).

4. For at least one criterion the following holds true: small differences of evaluations are not significant in terms of preferences, while the accumulation of several small differences may become significant. This requires the introduction of discrimination thresholds (indifference and preference) which leads to a preference structure with a comprehensive intransitive indifference binary relation.

In this paper, we introduce fuzzy Electre method. We suppose that the weights of attributes and the rates of options with respect to attributes are trapezoid fuzzy numbers.

In Section 2, we introduce the problem and notations. Section 3 and 4 develops Fuzzy Electre method. In Section 5, a numerical example is presented. Finally, Section 6 is devoted to the concluding remarks.

2. Problem Description and Notations

Electre is one of the multi-criteria decision making methods. This method concentrates on outranking relations between alternatives to find a subset of partially dominant alternatives. In this paper, we develop this method in a fuzzy situation which both weights of the criteria and ratings of the alternatives are expressed in trapezoid fuzzy numbers.

The parameters of this problem are as follows:

\( A_j \): Attribute \( j \)

\( O_i \): Option (alternative) \( i \)

\( n \): Number of attributes

\( m \): Number of options

\( w_j \): Weight of attribute \( j \)

\( c_{ke} \): Degree of concordance of relation “alternative \( k \) is at least as good as alternative \( e \)”

\( d_{ke} \): Degree of discordance of relation “alternative \( k \) is at least as good as alternative \( e \)”

To be more familiar with the definition of \( c_{ke} \) and \( d_{ke} \), note that:

\( c_{ke} \) informs us about the strength of preference for the alternative \( k \), with respect to the alternative \( e \) when considering all the criteria, while,

\( d_{ke} \) informs us about the evidences \textit{against} the preference of alternative \( k \) with respect to alternative \( e \) under the consideration of all of the criteria.

3. Fuzzy Electre Method

Suppose that the decision matrix is as follows:
\[
\begin{bmatrix}
A_1 & A_2 & \cdots & A_j & \cdots & A_n \\
O_1 & & & & & \\
O_2 & & & & & \\
\vdots & & & & & \\
O_t & & \cdots & \tilde{x}_{ij} & & \\
\vdots & & & & & \\
O_m & & & & & \\
\end{bmatrix}
\]

In this problem, both \( \tilde{x}_{ij} \) s and the weights of the attributes are trapezoid fuzzy numbers. So, we have:

\[
\tilde{x}_{ij} = (a_{ij}, b_{ij}, c_{ij}, d_{ij}) 
\]

\[
\tilde{\omega}_j = (\alpha_j, \beta_j, \gamma_j, \delta_j) 
\]

We suppose that all the fuzzy numbers are positive. In other words, for each trapezoid fuzzy number \( \tilde{\nu} = (\nu_1, \nu_2, \nu_3, \nu_4) \), we have:

\[
\nu_4 \geq \nu_3 \geq \nu_2 \geq \nu_1 \geq 0
\]

The fuzzy Electre method includes the following steps:

Step1. Smooth the range of fuzzy elements in every attribute using the following expression:

\[
\tilde{r}_{ij} = \begin{cases} 
\frac{a_{ij} + b_{ij} + c_{ij} + d_{ij}}{4d_j} & \text{for the positive attributes} \\
\frac{-a_{ij} - b_{ij} - c_{ij} - d_{ij}}{4d_j} & \text{for the negative attributes}
\end{cases}
\]

where;

\[
d_j^* = \max_i \{d_{ij}\}
\]

In expression 3, positive attributes are the attributes which the decision maker wants to maximize them (profit-like attributes), while negative attributes are the attributes which the decision maker tries to minimize them (cost-like attributes).

So, the smoothed decision matrix, which is called normalized decision matrix, would be as follows:

\[
R = \begin{bmatrix}
\tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1n} \\
\tilde{r}_{21} & \tilde{r}_{22} & \cdots & \tilde{r}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{r}_{m1} & \tilde{r}_{m2} & \cdots & \tilde{r}_{mn}
\end{bmatrix}
\]

Step2. Multiply each column of matrix R by the weight of its corresponding attribute.
\[ \tilde{v}_j = \tilde{r}_j \otimes \tilde{w}_j = (a_y \alpha_j, b_y \beta_j, c_y \gamma_j, d_y \delta_j) \] (5)

where, \( \tilde{r}_y = (a_y, b_y, c_y, d_y) \) and \( \tilde{w}_j = (\alpha_j, \beta_j, \gamma_j, \delta_j) \) are the elements of matrix \( R \) and the weights of attributes, respectively.

Step3. For each pair of options \( e \) and \( k \), compose the following sets:

\[ C_{ke} = \{ j \mid \tilde{v}_{kj} \geq \tilde{v}_{ej} \} \] (6)
\[ D_{ke} = \{ j \mid \tilde{v}_{kj} < \tilde{v}_{ej} \} \] (7)

Note that, for comparing two fuzzy numbers \( \tilde{x} \) and \( \tilde{y} \), we have:

\[ \text{IF} \{ M(\tilde{x}) > M(\tilde{y}) \ \text{OR} \ \{ M(\tilde{x}) = M(\tilde{y}) \ \text{AND} \ \sigma^-(\tilde{x}) < \sigma^-(\tilde{y}) \} \} \]
\[ \text{THEN} \ \tilde{x} \succ \tilde{y} \]

where, \( M(\tilde{x}) \) is the mean of \( \tilde{x} \) and \( \sigma^-(\tilde{x}) \) is its negative standard deviation. Especially, for a trapezoid fuzzy number \( \tilde{x} \), we have:

\[
\begin{align*}
\text{Degree of possibility} \\
\begin{array}{c}
\text{Value of } x \\
\hline
a & b & c & d
\end{array}
\end{align*}
\]

Figure 1: A trapezoid fuzzy number

\[
M(\tilde{x}) = \frac{c^2 + 3d^2 + cd - a^2 - b^2 - ab}{3(c + d - a - b)}
\] (9)

and;

\[
\sigma^-(\tilde{x}) = \sqrt{E[(x - M)^2]}
\] (10)

where;

\[
(x - M) = \begin{cases} 
 x - M & : x \leq M \\
 0 & : x > M 
\end{cases}
\] (11)

By substituting \( M(\tilde{x}) \) in (10), we have:

\[
\sigma^-(\tilde{x}) = \sqrt{ \int_a^b \left[ \frac{x - M}{b - a} (x - M)^2 \right] dx + \int_b^d \left[ (x - M)^2 \right] dx }
\] (12)

\[
= \left[ \frac{1}{2} \left( \frac{b^4}{4} - (2M + a) \frac{b^3}{3} + (M^2 + 2Ma) \frac{b^2}{2} - \frac{a^4}{4} \right) \\
+ (2M + a) \left( \frac{a^3}{3} - (M^2 + 2Ma) \frac{a^2}{2} + M^3 - Mb^2 + Mb^2 - \frac{b^3}{3} \right) \right] ^{\frac{1}{2}}
\]

Step4. Compose concordance matrix \( C \) and discordance matrix \( D \) for all of the relations “alternative \( k \) is at least as good as alternative \( e \)” as follows:
\[
c_{k\epsilon} = \frac{\sum_{j=1}^{n} M(\tilde{v}_{ij})}{\sum_{j=1}^{n} M(\tilde{v}_{ij})} \quad (13)
\]

and;

\[
d_{k\epsilon} = \frac{\max_{j \in D_{k\epsilon}} [M(\tilde{v}_{ij}) - M(\tilde{v}_{ij})]}{\max_{j=1,n} [M(\tilde{v}_{ij}) - M(\tilde{v}_{ij})]} \quad (14)
\]

where, \( M(.) \) is the mean function of fuzzy number.

Step 5. Compose the final dominance matrix \( (F) \) as follows:

5.1) Let

\[
\bar{c} = \sum_{k=1}^{m} \sum_{\epsilon=1}^{m} \frac{c_{k\epsilon}}{m(m-1)}
\]

5.2) Let

\[
\bar{d} = \sum_{k=1}^{m} \sum_{\epsilon=1}^{m} \frac{d_{k\epsilon}}{m(m-1)}
\]

5.3) The final dominance matrix would be as follows:

\[
f_{i\epsilon} = \begin{cases} 
1 & : c_{i\epsilon} \geq \bar{c} \text{ and } d_{i\epsilon} \leq \bar{d} \\
0 & : \text{Otherwise}
\end{cases}
\]

Step 6. Eliminate options which are significantly dominated.

Unfortunately in step 6, except the completely dominated options, identifying that which other options can be dominated is somehow vague. To solve this problem, one way is using the following measures:

\[
c_{k} = \sum_{\epsilon=1}^{m} c_{k\epsilon} - \sum_{\epsilon=1}^{m} c_{\epsilon k} \quad (18)
\]

\[
d_{k} = \sum_{\epsilon=1}^{m} d_{k\epsilon} - \sum_{\epsilon=1}^{m} d_{\epsilon k} \quad (19)
\]

where, options (alternatives) which have larger \( c_{k} \) and smaller \( d_{k} \) are more preferable.

4. Numerical Example

Suppose that the initial decision matrix and the weights of the attributes are as follows:
\[ \vec{w}_1 = (1, 2, 3, 4); \vec{w}_2 = (2, 5, 6, 8); \vec{w}_3 = (3, 4, 6, 7) \]

Furthermore, suppose that the second attribute is negative.

Solution:

The weighted normalized decision matrix would be as follows:

\[
V = \begin{bmatrix}
(0.11, 0.33, 0.55, 0.89) & (-1.00, -0.89, -0.78, -0.22) & (0.25, 0.37, 0.50, 0.62) \\
(0.11, 0.55, 0.78, 0.89) & (-0.89, -0.78, -0.67, -0.55) & (0.25, 0.75, 0.87, 1.00) \\
(0.55, 0.78, 0.89, 1.00) & (-0.89, -0.78, -0.33, -0.11) & (0.12, 0.37, 0.62, 0.75)
\end{bmatrix}
\]

The mean values of the elements of matrix \( V \) are shown in matrix \( M \):

\[
M = \begin{bmatrix}
A_1 & A_2 & A_3 \\
O_1 & 0.47 & -0.69 & 0.43 \\
O_2 & 0.57 & -0.72 & 0.69 \\
O_3 & 0.80 & -0.52 & 0.46
\end{bmatrix}
\]

The negative standard deviation of the elements of matrix \( V \) are shown in matrix \( \sigma^- \):

\[
\sigma^- = \begin{bmatrix}
A_1 & A_2 & A_3 \\
O_1 & 0.14 & 0.25 & 0.12 \\
O_2 & 0.19 & 0.26 & 0.29 \\
O_3 & 0.32 & 0.20 & 0.14
\end{bmatrix}
\]

The dominance and inferiority sets for each pair of options are:

- \( C_{12} = \{2\}; \ D_{12} = \{1, 3\}; C_{13} = \{\}; D_{13} = \{1, 2, 3\}; C_{23} = \{3\}; D_{23} = \{1, 2\} \)
- \( C_{21} = \{1, 3\}; D_{21} = \{2\}; C_{31} = \{1, 2, 3\}; D_{31} = \{\}; C_{32} = \{1, 2\}; D_{32} = \{3\} \)

The dominance matrix (\( C \)) and the inferiority matrix (\( D \)) are as follows:

\[
C = \begin{bmatrix}
O_1 & 0.39 & 0 \\
O_2 & 0.61 & -0.42 \\
O_3 & 1 & 0.58
\end{bmatrix} \quad D = \begin{bmatrix}
O_1 & -1 & 1 \\
O_2 & 0.12 & -1 \\
O_3 & 0 & 1
\end{bmatrix}
\]

Furthermore, we have:

\[ \bar{c} = 0.50 \; ; \; \bar{d} = 0.69 \]

So, the final dominance matrix would be:
As shown in matrix $F$, option $O_1$ is dominated by options $O_2$ and $O_3$, so this option is eliminated from further evaluation. For obtaining a more clear view about the dominance between $O_2$ and $O_3$, we use expressions 18 and 19. Thus, we have:

c_2 = 0.06; d_2 = -0.88

c_3 = 1.16; d_3 = -1

Since $c_3 \geq c_2$ and $d_3 < d_2$, we can conclude that option 2 is partially dominated by option 3.

5. CONCLUDING REMARKS

In some real-world situations, because of incomplete or unavailable information, the data are imprecise and expressed as fuzzy numbers. Electre methods are among the most popular MADM methods with a wide scope of application, like: classification of environmental impacts [12], facility layout problem [13], material selection [14], outsourcing (vendor selection) problem [15], etc. Despite the importance of Electre methods and the popularity of fuzzy environments, the fuzzy version of Electre methods was not published by the time that we submitted this paper. Hence, in this paper we developed Electre method for decision making in fuzzy environments. Our developed method is essentially based on Electrel with some small modifications. We suppose that the ratings of options and the weights of attributes are trapezoid fuzzy numbers. Since each triangle is an especial kind of trapezoid, our discussions can also be used for triangular fuzzy numbers.

REFERENCES


Driving Method Study and its Implementation for TFT-LCD with High Gray Scale

RAN Feng\textsuperscript{a,b,*}, XU Meihua\textsuperscript{a,c}, CEHN Zhangjin\textsuperscript{a,b}

\textsuperscript{a} Microelectronic Research and Development Center, Shanghai University
\textsuperscript{b} Key Laboratory of Advanced Displays and system Application, Ministry of Education, Shanghai University
\textsuperscript{c} School of Mechatronical Engineering and Automation, Shanghai University
Shanghai, 200072, China

\textbf{ABSTRACT}

This paper presents an efficient pulse width modulation (EPWM) drive method, which is used to implement the driving controller for TFT-LCD with high gray scale. To solve the time redundancy issues in the traditional pulse width modulation (TPWM) drive method, the proposed method divides the display panel into several sub-partitions and scans the sub-partitions alternately, and the gray data and vanishing signals are transmitted with separate wires in driving controllers. It’s theoretically proved that the drive efficiency can achieve 100% for any gray scale display systems with EPWM, and the implementation also shows that, comparing with the TPWM method, the minimum clock frequency in the TFT-LCD systems can greatly reduced with the proposed method.

\section{1. Introduction}

Because of their good overall performance, the thin film transistor liquid crystal displays (TFT-LCDs) have been becoming the mainstream products on the flat-panel display market. For TV applications, customers want not only the large screen size but also natural colours. So, larger screen size and higher definition have always been the goal manufacturers pursuing. However, the improvement of the screen size and definition is restricted by many factors, one important of which is the gray scale modulation method.

There are two voltage driving methods to realize the gray scale modulation of TFT-LCDs. Pulse amplitude modulation (PAM) method modulates the voltage amplitude to generate different gray scales. Pulse width modulation (PWM) method modulates the lighting time with a constant voltage through the pixels.\cite{1, 2, 3}

In order to improve the gray scale, the PAM needs high-accurate voltage amplitude scale while the PWM needs high scan frequency. However, it is difficult to realize the high-accurate voltage digital-to-analogue convector (DAC) and reach high system drive clock.\cite{4, 5}

In this paper, section II analyses the time redundancy issue existing in the Traditional Pulse Width Modulation (TPWM) drive method. Section III presents the Efficient Pulse Width Modulation (EPWM) drive method and proves that the drive efficiency of the EPWM drive method will be improved greatly. In section IV, the implementation of EPWM and the comparison between EPWM and other PWM drive methods are presented.

\section{2. Time Redundancy Issue of Traditional Pulse Width Modulation Method}

PWM (Pulse Width Modulation) drive method can be realized through several different forms. The time redundancy issues existing in the sub-frame modulation method, as an example, will be discussed.

The gray scale with the sub-frame modulation can be achieved by controlling the frame valid driving time according to the bit power of the pixel gray data. The gray value of a pixel is shown in the formula (1).
For a TFT-LCD with 8 levels gray scale, sub-frame modulation method divides a scan cycle into 8 time slices (or 8 sub frames). Then, the pixels can achieve different gray scales by turning on or turning off pixel voltage during the corresponding time slices shown in Table 1.

Table 1: Sub-frame modulation of 8 levels gray scale

<table>
<thead>
<tr>
<th>Nts</th>
<th>G/B</th>
<th>B0</th>
<th>B1</th>
<th>*</th>
<th>B2</th>
<th>*</th>
<th>*</th>
<th>Bx</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

In Table 1, Nts represents the sequence number of the time slices. G represents the gray value and B represents the bit of the gray data. Bx represents the vanishing cycle.

From the table 1, we can find that the pixel voltage has the same state, turning on or turning off, at the 2nd and 3rd time slices. This means that the controller only needs to transmit a bit of gray data to the pixel at the 2nd time slice. The same situation occurs at the 4th, 5th, 6th and 7th time slices. The controller transmits available data, including gray data and vanishing signal, only at the 1st, 2nd, 4th and 8th time slices. We call these time slices available scan cycles, while other time slices are waiting time, which is called redundant time. Then the drive efficiency of sub-frame modulation is:

\[ E_d = \frac{N_d}{N_{all}} \times 100\% \]

Where \( N_d \) is the number of available scan cycles, and \( N_{all} \) is the number of total scan cycles.

It can be seen that the drive efficiency of TPWM is rather low. The minimum clock frequency will be higher and higher with the continuous increase of gray scale, which makes the display system hard to implement in practice.

3. Introduction of Efficient Pulse Width Modulation

To take full use of the redundant time, the TFT-LCD panel is divided into \( 2^n-1 \) sub-partitions [6, 7, 8] while any sub-partition can get the same target gray scale. The variable \( n \) is the number of bit of gray scale. One sub-partition is scanned in the redundant time of the other sub-partitions. By doing this, the drive efficiency will be improved greatly. Meanwhile, gray data is separated from vanishing signal in transmission process. Using this method, the scanning time of vanishing signal in TPWM method can be decreased. As a result, the clock frequency of the display system will be further reduced.

Our goal is to calculate the maximum sub-partition number (SN) and seek the proper drive sequence in order to maximize the scanning efficiency to attain about 100%. We will take the 4 levels gray scale and 8 levels gray scale as examples to illustrate how the controller scanning in different sub-partitions alternately.

3.1. Drive Sequence of 4 Levels Gray Scale

Our goal is to calculate the maximum sub-partition number (SN) and seek the proper drive sequence in order to maximize the scanning efficiency to attain about 100%. We will take the 4 levels gray scale and 8 levels gray scale as examples to illustrate how the controller scanning in different sub-partitions alternately.

For 4 levels gray scale display system, the TFT-LCD panel will be divided into two sub-partitions, labeled S1 and S2. Then they will be scanned alternately according to the order shown in Table 2.
Controller sends the available data in the following order: $B_1 (S_1) \rightarrow B_1 (S_2) \rightarrow B_0 (S_1) \rightarrow B_1 (S_2)$. A matter worthy of note is that the vanishing signal and gray data in 4<sup>th</sup> time slice will not conflict. This is because they are transmitted by two separate data lines. All of the five time slices are available scan cycle, so the drive efficiency is 100%.

### 3.2. Drive Sequence of 8 Levels Gray Scale

While the gray scale is 8 levels, the TFT-LCD panel will be divided into four sub-partitions, labeled $S_1$, $S_2$, $S_3$, and $S_4$. The drive sequence of 8 levels gray scale is shown in Table.3.

<table>
<thead>
<tr>
<th>$N_n$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
<th>$S_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$B_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$B_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$B_0$</td>
<td></td>
<td>$B_0$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$B_1$</td>
<td></td>
<td></td>
<td>$B_2$</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>$B_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$B_0$</td>
<td>$B_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$B_1$</td>
<td>$B_0$</td>
<td>$B_1$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$B_1$</td>
<td>$B_0$</td>
<td>$B_1$</td>
<td>$B_2$</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>$B_1$</td>
<td>$B_0$</td>
</tr>
<tr>
<td>10</td>
<td>$B_1$</td>
<td></td>
<td></td>
<td>$B_0$</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>$B_0$</td>
</tr>
<tr>
<td>12</td>
<td>$B_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be seen clearly that the drive efficiency is still 100%, all of the thirteen time slices are available scan cycles. It will be discovered that the drive sequence of 8 levels gray scale can be derived from that of 4 levels gray scale when vanishing signals are not considered. This will be verified in the following section. It’s emphasized that additional vanishing signals are inserted to ensure the pixels voltage turning on for a valid time. For example, a vanishing signal is send to the sub-partition $S_3$ at the 7<sup>th</sup> time slice to ensure the bit power of $B_2$ maintaining four time slices. This is different from TPWM method, in which the vanishing signals are scanned lastly.

### 4. Theoretical Derivation of Efficient Pulse Width Modulation

We transform the drive sequence of 4 levels gray scale into matrix form, which is called drive matrix, shown in formula (3).

$$ M_4 = \begin{bmatrix} B_1 & & & \\ & B_1 & & \\ & & B_0 & & \\ & & & B_0 & \end{bmatrix}_{4 \times 2} $$

Similarly, the drive sequence of 8 levels gray scale can be expressed in matrix form in formula (4).
Where, $E_4$ equals to

$$
\begin{bmatrix}
B_2 \\
B_1 \\
B_0 \\
B_2
\end{bmatrix}
$$

which is called identity matrix. $0_{4\times2}$ is a zero matrix with 4 rows and 2 columns. All of the omitted elements in matrix $M_2$, $M_3$, $E_4$ are zero.

It can be discovered that the drive matrix $M_3$ can be derived from drive matrix $M_2$, and $M_2$, $M_3$ have some characteristics in common. Both matrix $M_2$ and $M_3$ have only one non-zero element every row, meaning only one available data is transmitted at every time slice. The number of non-zero elements in every column is 2 and 3 respectively, corresponding to the number of bit of gray scale.

A drive matrix must satisfy three conditions as follows.

1. The number of rows is $n \cdot 2^{n-1}$, and the number of columns is $2^{n-1}$. Where $n$ is an integer no less than 1, corresponding to the number of bit of gray scale.

2. There is only one non-zero element every row.

3. The number of non-zero elements in each column is $n$.

If a matrix $M_{n-1}$ is the drive matrix of $2^{n-1}$ levels gray scale, we can prove the matrix $M_n$ shown in formula (5) is the drive matrix of $2^n$ levels gray scale.

$$
M_n =
\begin{bmatrix}
\begin{array}{c@{}c@{}c}
E_{2^{n-1}} & 0 & \vdots \\
\vdots & \ddots & \vdots \\
0 & \ldots & M_{n-1}
\end{array}
\end{bmatrix}
$$

$M_{n-1}$ is the drive matrix of $2^{n-1}$ levels gray scale, which means there is only one non-zero element every row and there are $n-1$ non-zero elements every column. As $E$ is an identity matrix, it can be discovered easily that the matrix $M_n$ also meets the three conditions the drive matrix required. So, $M_n$ is the drive matrix of $2n$ levels gray scale.

The row number of matrix $M_n$ is $n \cdot 2^{n-1}$ and the number of non-zero elements in $M_n$ is $n \cdot 2^{n-1}$. If the drive sequence of display device is expressed as $M_n$, the controller will transmit $n \cdot 2^{n-1}$ available data at $n \cdot 2^{n-1}$ time slices. This implies the drive efficiency is 100%.
We can get the following conclusions through the analysis. For a n-bit gray scale display system, if it divides the whole TFT-LCD panel into $2^n - 1$ sub-partitions and employs the two-lines drive structure and the drive sequence expressed by $M_n$, the drive efficiency will attain 100%.

5. Implementation of Efficient Pulse Width Modulation

The implementation architecture of is shown in Figure.1.

In figure.1, the image data converter module decodes the image signal from micro processor unit (MPU) or digital visual interface (DVI), and writes the output RGB data to RAM. The controller of EPWM module is the key module in the display system. It not only controls the drive sequence of sub-partitions in accordance with drive matrix, but also generates the vanishing signals which determine the illumination time of the pixel together with gray data. The row driver IC and column driver IC generate gate and source driving voltage orderly, under the control of EPWM controller.

Clock frequency is one of the standards to measure the feasibility of a drive method. The TPWM drive method is impossible to realize, just because the clock frequency is too high to attain.

In the PWM drive method with no sub-partition division, the minimum clock frequency is required as:

$$f_{clk} = \text{refresh frequency} \times \text{row resolution} \times \text{column resolution} \times \text{gray scale levels}$$

(6)

In the PWM drive method with sub-partition division such as the Fractal Scanning Drive Method (FSDM)[7], the clock frequency is required as:

$$f_{clk} = \text{refresh frequency} \times \text{row resolution} \times S \times \text{pixels number per sub-partition}$$

(7)

Where, $S$ is the cycle number of a complete scan. It includes the scan of all gray data and vanishing signals of one pixel in each sub-partition.

In the practical interface between controller and display panel, the data width transmitted every clock is 8 bits or 16 bits. It’s assumed that the data width is 8 bits, then the actual clock frequency is

$$f_{act} = f_{clk} / 8$$

(8)

Then, for a 1024×768 dots RGB TFT-LCD, we can analogically calculate the drive efficiency and minimum clock of different gray scales with different drive methods, while the refresh rate is 60 Hz and data width is 8 bits. It is shown in table.4.

<table>
<thead>
<tr>
<th>Gray scale</th>
<th>Drive method</th>
<th>Drive efficiency</th>
<th>Minimum clock (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>TPWM</td>
<td>10.94%</td>
<td>377.49</td>
</tr>
<tr>
<td></td>
<td>FSDM</td>
<td>100%</td>
<td>41.29</td>
</tr>
</tbody>
</table>
From table 4, it can be seen TPWM is most inefficient and the minimum clock required is highest. When the gray scale is 1024 levels, the minimum clock frequency reaches 6057.49 MHz, which is impossible for the practical TFT-LCD display systems. Both of the FSDM and the proposed EPWM method can obtain 100% drive efficiency by making full use of redundant time, while the minimum clock frequency is far lower in comparison with that of TPWM. Since EPWM controller transmits the gray data and vanishing signals separately, the minimum clock frequency required is lower than that of the FSDM. For the 1024 levels gray scale display system, the minimum clock frequency with EPWM is only 58.99 MHz.

6. SUMMARY

The design calculation and system realization indicate that the proposed method can easily realize the high gray scale drive for TFT-LCD. With sub-partition division and two-line drive structure, the proposed method can obviously improve the drive efficiency and reduce the minimum clock. As a result, the realization difficulty of the high gray scale system is reduced.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the National Natural Science Foundation of China for providing financial support for this work under Grant No.60773081 and Grant No.60777018, and also to acknowledge the financial support by Shanghai Science and Technology Department under Grant No.077062008 and Grant No.08706201800.

REFERENCES

A Novel OLED Controller with Fractal Scan Scheme

XU Meihua\textsuperscript{a,b,*}, RAN Feng\textsuperscript{b,c}, and CEHN Zhangjin\textsuperscript{b,c}

\textsuperscript{a}School of Mechatronical Engineering and Automation, Shanghai University
\textsuperscript{b}Key Laboratory of Advanced Displays and system Application, Ministry of Education, Shanghai University
\textsuperscript{c}Microelectronic Research and Development Center, Shanghai University
Shanghai, 200072, China

\textbf{ABSTRACT}

Based on the mathematical model of the optimal scan architecture of grayscale imaging, this paper presents the design and logic implementation of a novel OLED greyscale controller with the fractal scan scheme to efficiently increase the scan utilization and imaging quality. Through the exploration of the sub-space code sequences and bit code sequences for different gray levels, we firstly complete the design and implementation of the parameterized fractal scanning IP core for various gray levels, which can be used to every flat panel display controller. And then we complete the design of a novel OLED grey scale controller that the fractal scanning IP core described with Verilog language is embedded in the FPGA hardware frame, which can efficiently increase the imaging gray levels and speed up frame frequency of display systems. Serial applications and tests indicate that it can obtain 100% scan utilization and a significant decrease of the system clock frequency (a multiple of 29 for 256 gray levels), comparing with traditional scan methods. With the proposed method there is no need to using high-speed IC circuits to realize the high articulation and resolution display of video images. Thus, the paper provides a new approach and an engineering way to solve one of the pressing problems of high resolution flat panel display technology.

1. INTRODUCTION

The Flat Panel Display (FPD), including TFT-LCD(Thin Film Transistor-Liquid Crystal Display), OLED(Organic Light Emitting Display), etc., that has been rapidly replacing bulky CRT based traditional displays plays an important role in exploring the progress of modern IT devices\cite{1,2}. As the high resolution flat panel display technology for large display size has been one of the key research fields, one of the bottlenecks which impede the technology development from current small display-size flat panel displays to high resolution large display-size ones are low grayscale scan utilization and efficiency. Xu et al. \cite{3, 4} proposed an innovative fractal scan structure which significantly increases scan efficiency. In their paper, they studied different scan schemes and constructed a mathematical model through the analysis of the fractal characteristics of space-time mapping topology. The theoretic model proved an optimal scan structure with maximum scan utilization.

Based on the aforementioned mathematical model\cite{5}, this paper presents the design and logic implementation of the fractal scan algorithm as a fractal scan intellectual property (IP) core for various gray levels, and then presents the system design and realization for OLED grayscale controller to testify the feasibility of using a fractal scanning IP core in the industrial applications, and the contribution for improving the parameters such as gray scale, frame frequency in FPD system. Compared with traditional scan methods, simulation and real application show 100% scan utilization and a significant decrease of system clock frequency (a multiple of 29 for 256 gray levels). Thus, the paper provides a new light to solving one of the pressing problems of high resolution FPD technology.

2. IP CORE WITH FRACTAL SCAN SCHEME FOR FPD

2.1. Fractal Topology for Optimal Scan

* Corresponding author: Tel.: (8621) 56331932; Fax: (8621) 56331632; E-mail: mhxu@shu.edu.cn
A key to the research of grayscale image scan algorithm is to convert the grayscale storage matrix into pixel lightening time. Hence, a space-time mapping topology for grayscale scan can be constructed at first. The grayscale matrix in space is then mapped to a grayscale scan space-time plane. Based on this mapping architecture, the optimal scan structure is explored. Since the paper is mainly focused on the logic implementation of the optimal scan algorithm and its application, interested readers can refer to the literature [3, 4, 5] for more details on different scan structures and methods.

Consider a scan line and rotate it by 90 degrees backward. Then the line and the time axis form a time-space plane, in which the horizontal axis represents the pixels and the vertical axis the grayscale scan time. For n-bit grayscale (n bits in the binary representation of the gray level), the entire time-space plane is partitioned vertically into Mn smaller planes (sub-spaces), each being of the same width. In each sub-space, there are n scan bits (scan for n times). For simplicity of discussion, denote Hk as the grayscale weight of the kth scan bit, E and T as the sum of weight of each scan bit, and the total scan time which is the sum of E and the blanking interval. As the plane is divided into 2n-1 sub-spaces, the optimal scan structure makes use of the interval between two scans of the same pixel (in the same column) to scan the pixels in other sub-spaces. The new scan scheme, according to the fractal characteristics of the structure, takes full advantage of the time interval and achieves 100% scan utilization.

Then rotate the time-space plane by 90 degrees anticlockwise for custom, namely the horizontal axis represents the grayscale scan time and the vertical axis the pixels. Figure 1 illustrates the zigzag fractal topology. The horizontal axis represents scan time, 40 scan time for every frame, and the vertical axis represents sub-spaces whose number is decided by the equation: Mn=8. The geometrical figures (■●◆▲) respectively denote the position and weight of the sub-space scanned currently. The rectangles beginning at black figures express the stable display time as soon as the corresponding sub-space received the scan data. The rectangle length is directly proportional to the weighted value. The white pentagrams show the last blanking disposal for every frame.

Each of the eight sub-spaces is scanned 5 times in total at the time (horizontal) axis. The scan distance for the first four scans is among 1 (H1=1), 3 (H2=3), 8 (H3=8) and 20 (H4=20), which corresponds to the grayscale weight. At each time unit, only one sub-space is scanned. The last eight scans (t32 to t39) deal with blanking.

Connecting every geometrical figure in term of time, the “Z” pattern emerges as the self-similar fractal topology.

2.2. Architecture of Fractal IP Core

The fractal scan module is shown in Figure 2. Each signal is defined in Table 1. The function of the module is to implement the fractal algorithm in digital logic. In the figure, NB is the bit width of the bit code and NS the bit width of the sub-space code. NB is determined by the gray levels and NS by the scan scheme. Since there are 2n-1 sub-spaces for n-bit grayscale, and NS = n-1. At the positive edge of the clock when GetNext is asserted (GetNext = 1), the module outputs current bit code Bit[NB-1:0], sub-space code Segment [NS-1:0], and the Hidden signal to control row/column addresses and gray value.

---

Figure 1: Fractal topology of optimal scan structure for 4-bit grayscale

Figure 2: Fractal scan module
Table 1: Interface signal definition for fractal scan kernel module

<table>
<thead>
<tr>
<th>Signal</th>
<th>Name</th>
<th>Bit width</th>
<th>Direction</th>
<th>Commentary</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLK</td>
<td>Clock</td>
<td></td>
<td>IN</td>
<td>Clock</td>
</tr>
<tr>
<td>rstn</td>
<td>Reset</td>
<td></td>
<td>IN</td>
<td>Reset. Valid at low voltage level</td>
</tr>
<tr>
<td>GetNext</td>
<td>Next</td>
<td></td>
<td>IN</td>
<td>Output next scan sequence, Valid at high voltage</td>
</tr>
<tr>
<td>Segment [NS-1:0]</td>
<td>Sub-space</td>
<td>NS</td>
<td>OUT</td>
<td>Sub-space code in sequence</td>
</tr>
<tr>
<td>Bit [NB-1:0]</td>
<td>Bit code</td>
<td>NB</td>
<td>OUT</td>
<td>Bit code in sequence</td>
</tr>
<tr>
<td>Hidden</td>
<td>Blanking</td>
<td></td>
<td>OUT</td>
<td>blanking code in sequence, Valid at high voltage</td>
</tr>
<tr>
<td>Last</td>
<td>Last</td>
<td></td>
<td>OUT</td>
<td>The last code during blanking, Valid at high</td>
</tr>
</tbody>
</table>

Before designing the logic circuit, we list weight sequences for every sub-space in terms of the time axis, as shown in weight sequences column of Table 2. All the weight sequence for every sub-space is an array including 1, 3, 8, 20, and possesses a kind of symmetric structure. As 4 bits in the binary representation is need to express 16 gray levels, the bit code sequence can be got from weight one by Bit0 corresponding weight 1, Bit1 corresponding weight 3, Bit2 corresponding weight 8 and Bit3 corresponding weight 20, which is listed in right column of Table 2.

And then listing sub-space and bit code sequences for every scan time, for example, during time 0, it is corresponding to the 0th sub-space, weight 20, bit code 3 and blanking code 0. Similarly, the sub-space code, bit code and blanking code will be calculated for other scan time, which are shown in figure 1. Apparently, the figure 1 and the Table 2 can be deduced each other.

To avoid circuit hazards, the eight sub-spaces are encoded with the Gray code which has the property that when advancing from one index to the next adjacent index, only a single bit changes value. In actual hardware design, quasi-Gray code is adopted.

Table 2: Weight and bit sequences (n = 4)

<table>
<thead>
<tr>
<th>Sub-space encode</th>
<th>Weight sequences</th>
<th>Bit code sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20 8 3 1</td>
<td>3 2 1 0</td>
</tr>
<tr>
<td>1</td>
<td>20 8 1 3</td>
<td>3 2 0 1</td>
</tr>
<tr>
<td>3</td>
<td>20 3 1 8</td>
<td>3 1 0 2</td>
</tr>
<tr>
<td>2</td>
<td>20 1 3 8</td>
<td>3 0 1 2</td>
</tr>
<tr>
<td>6</td>
<td>8 3 1 20</td>
<td>2 1 0 3</td>
</tr>
<tr>
<td>7</td>
<td>8 1 3 20</td>
<td>2 0 1 3</td>
</tr>
<tr>
<td>5</td>
<td>3 1 8 20</td>
<td>1 0 2 3</td>
</tr>
<tr>
<td>4</td>
<td>1 3 8 20</td>
<td>0 1 2 3</td>
</tr>
</tbody>
</table>

Figure 2: Block diagram of the fractal scan module

2.3. Fractal IP Core Design

To meet the requirement of different grayscale scans, we design a parameterized module which covers from a minimum of 4 to a maximum of 65535 gray levels.

The block diagram of the IP core is illustrated in Figure 3. In this figure, Gray Location Code and Parity Flag generate the next gray code for the current sub-space code Segment. BST is the current state encoding in the state transition diagram for bit code sequence. The control modules, Count and Finish, record the current counting state. Bit Code Parity Count K records the number of occurrences for each state.

Next state code, NextBST, is the code generated for the next clock cycle. Note that the current state is the combination of BST, current count state and K to generate, for the next sub-space, bit code select signal Sel, transition code D as well as the bit code (Bit), blanking code (Hidden) and cycle done code (Last). The XOR logic and the multiplexer output the next sub-space code NextSeg.
For a parameterized IP core, we can obtain simulation waveforms for various grayscales. Figure 4 shows the waveform for 16-bit grayscale ($n=4$, $NB=2$) in the Active HDL™ software environment. At the 10MHz clock frequency, the simulation waveform displays the code sequences Segment, Bit and Hidden as expected in Figure 1.

### 3. NOVEL OLED CONTROLLER EMBEDDED FRACTAL IP CORE

The fractal scanning controller is composed of fractal scanning control circuits (using Altera's EP1K50QC208), synchronous static memory (using ISSI’s IS61NP25636), four-channel fast Ethernet Transceiver (using Altima’s AC104-QF), four-channel filter transformer (using Pulse’s H1044), and the other peripheral interface circuits. Its module structure diagram is shown in Figure 5. The network ports management module receives signals from four RJ45 connectors one of which is selected as the image input channel. The main function of the network packet writing control module is to identify the network packets, cache and manage the data packets, and transform the gray scales of image packets.

As the key module of the entire scanning system, the memory management one designed with PIPELINE and NO-WAIT mode can make the chip have a complete data access at each clock. The efficiency of data bus is up to 100% with the access speed above 100MHz.
Main functions for the scanning control module are described as follows: The IP core is the parameterized fractal scanning module. The scanning sequence generator outputs the corresponding address and command of the next scanning period. The scanning reading control module accesses the memory according to the order sequence, and stores the data to the scanning buffer. The scanning sub-module reads the data from scanning buffer, and outputs it to the display panel at a constant speed. The memory management module takes corresponding control, coordination and handling to the external memory access.

3.1 MEMORY MANAGEMENT MODULE

Designing mode and its performance for the memory module directly influences other modules’ running functions. To achieve higher data throughput and simplify system design, the external memory selects static synchronous memory to simplify time sequences management and to reduce the access cycles [6], meanwhile, to improve the access performance of memory to the most extent especially when scanning discrete address leaps.

The memory access interface is shown in Figure 6. The letter “A” stands for address, D for data, GCLK for synchronous clock of FPGA and memory. CE represents the chip selection signal, WE for writing control signal and OE for reading control signal, all of which are valid at low voltage level. In addition, there are following signals generally included in memory: ADV is used as fresh command control signal, CKE as clock effective signal and BW as byte enable signal, which are connected to low voltage level to simplify the system design.

As a chief module in FPGA, the memory management module is required to adapt each memory access requirements, and keep the aggregate throughput as well, because each sub-module may need to access memory. The source data module, which wants to write data into memory, also needs to read from it for image comparison. The scanning module as well should output data to FPD panels from memory. Several scanning modules are required for a complex FPD system to scan different parts of the image.

3.2 GRAYSCALE TRANSFORM AND TRANSPOSE MODEL

There are two advantages to transform data with 256 gray levels into 4096 ones. One is to improve the display effect, and another is to obtain 256 gray levels from 4096 ones, which most adapt to the characteristic of diversified displays. The grayscale transform and transpose model is shown as Figure 7. As the package cache controller, the ntxch will write the network data to the package cache (ch1), and then, while the greyscale of data in packages have been transformed and transposed by a program, it will be written to another package cache (ch2), which can be directly sent to the memory.

In the model there are two caches which have different structure: ch1(256×24bit) & ch1 (256×32bit). Every cache is divided into two areas where writing and reading are separately carried out. The buffer controller is used to judge the status if the areas are empty or overflow, output operating codes and determine the start or pause operation of transform. The model selects one 8-bit color value among the red, green and blue ones after 24-bit colour value of the current pixel being read, and then gets the 12-bit grayscale value via looking up the transform table which is preset during FPGA design in advance. It can be modified by the control command word of network to obtain different grayscale...
transform effects. The 4-bit grayscale values selected from the 12-bit ones are sent to 4-channel SIPO respectively. After 32 time series input, the 4-channel 32-bit data will be received and then sent to cache ch2 via the multiplexer.

Figure 7: Grayscale transform and transpose

3.3. Fractal Scanning Control Module

The structure diagram of scan control module (as shown in Figure 8) is composed of five parts: fractal scan IP core sub-module, scan sequence generator sub-module, scan control sub-module, scan data cache sub-module and transmission control sub-module.

(1) Fractal Scan IP Core Sub-module

The fractal scan method of grayscale control for FPD system would make use of time redundancy to the greatest extent, and it has an optimal structure with highest scan efficiency [2]. The fractal scan IP core, implementing the arbitrary grayscale with hardware, is shown in Figure 3, and signal definition in Table 1.

Parameter $NB$, in Figure 2, standing for wide of bit code, depends on grayscale, $NS$, for bit-wide of subspace code, depends on scan method. Assuming $n$ for gray bits, gray level equals $2^n$, and $NB = \log_2 n$. Using proposed fractal scan method, the number of subspace is $2^{n-1}$, and $NS = n - 1$ [2]. The main function of the fractal scan core module is to achieve the logic implementation for fractal scan algorithms. At the time of not only rising edge for the signal CLK, but also high-level available for the GetNext signal, it outputs the current subspace code, bit code and blanking code, in the form of fractal to control the row-column driving address and OLED grayscale.

To avoid logical competition in hardware implementation, subspaces will be allocated dimensions in the form of Gray code (as 0, 1, 3, 2, 6, 7, 5, 4). When real scanning, in the direction of time axis, dimension of scanned subspace changes based on the special Quasi-Gray way.

As shown in Figure 2, in grayscale of 2048, there will be 1024 subspaces (Segment [9:0]), 11 bit codes (Bit [3:0]). The Hidden signal represents for blanking indication, the Last signal for accomplishment of scan once, and the GetNext for next scan code requiring.

(2) Scan Sequence Generator Sub-module
The control interface is designed based on fractal scan IP core (as shown in Figure 8). The iGetNext signal is a control signal requiring for next scan address, which is used to count for colour. Each colour of one pixel shares the same subspace and bit code. It will send the GetNext signal to the fractal scan IP core requiring for next subspace code when counted to three. When scanned to the last subspace code, the Last signal is available, and scan line will count (Line [1:0]) at the next GetNext signal available. The result of two counters (Column and Line) and the current code of scan IP core (Segment, Bit and Hidden) will be synchronously output at the next iGetNext signal available (iColumn, iLine, iSegment, iBit, iHidden), which could be supplied to generate memory access order for the scan control sub-module.

![Figure 8: Scan module structure](image)

(3) Scan Control and Cache Sub-modules

First, the cache controller should detect whether there is free scan cache space. If so, the scan command generator will require for access to memory management module, and wait for authorization, then, the command generator will continuously send the iGetNext signal to the scan sequence generator, change the address code obtained to row-column code according to the table, and further transfer it to memory access requirement code sent to memory management module. After fulfilling the corresponding requirements in memory access, the rstNew signal is available, and the data read from the memory will be written to the cache.

If the scan command generator finishes all the commands in one cache space, memory access will be stopped. At the same time, the cache controller indicates the full state of the cache, and wait for another cache space read free by transmission controller to launch the next cycle scan.

(4) Transmission Control Sub-module

According to the frequency division setting, the transmission mode of 1/2, 1/4 and 1/6 frequency division is adopted for the controller to master the counting of 8-bit address counter, as a result of address to read the scan cache for the data, which will generate the corresponding panel driving signals. Due to the gray generation mechanism, address counting must be uniform and continuous. Transmission control module does not matter whether the data from scan cache is expired. The transmission control sub-module should ensure is that once the cache space is read empty, it must be filled immediately.

4. SUMMARY

The OLED display systems have been designed and accomplished to validate application of the fractal scan IP core. The system is made up of the multimedia video card, the Ethernet video transmitter unit, the fractal scan controller and the OLED flat panel display. Using the fast Ethernet agreement technique to substitute the traditional difference data transfer method obtained the advantages such as fast speed, long distance data transmission and the high reliability of transceiver data [7]. The novel control mode that the fractal scan IP core described with Verilog language is embedded in the FPGA hardware frame can efficiently increase the imaging gray scales and quality in the OLED system.
Adopting parallel decoding joined with fractal scan technique can significantly raise the frame frequency of display system. All the scanning availabilities will be up to 100% and the scanning clock will be falling rapidly with proposed method and its performance listed in Table 3 has been testified in experiments in the condition of 100Hz frame frequency (every frame scanning 32 rows, every row including 512 pix), all the scanning availabilities with fractal scan mode can achieve 100% perfectly, demand for scanning clock will boost linearly, 256 gray levels with fractal scan mode can easily be implemented. Real application verifies that high resolution image display can be achieved in this system.

Table 3: Scan efficiency comparison between traditional mode and fractal mode

<table>
<thead>
<tr>
<th>Scan bit n</th>
<th>Gray level 2^n</th>
<th>Scan mode</th>
<th>Available scan (times)</th>
<th>Actual scan (times)</th>
<th>Scan efficiency</th>
<th>Scanning clock(MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
<td>Traditional mode</td>
<td>5</td>
<td>16</td>
<td>31.25%</td>
<td>9.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fractal mode</td>
<td>40</td>
<td>40</td>
<td>100.0%</td>
<td>3.07</td>
</tr>
<tr>
<td>7</td>
<td>128</td>
<td>Traditional mode</td>
<td>8</td>
<td>128</td>
<td>6.25%</td>
<td>78.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fractal mode</td>
<td>512</td>
<td>512</td>
<td>100.0%</td>
<td>4.92</td>
</tr>
<tr>
<td>8</td>
<td>256</td>
<td>Traditional mode</td>
<td>9</td>
<td>256</td>
<td>3.52%</td>
<td>157.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fractal mode</td>
<td>1024</td>
<td>128</td>
<td>1152</td>
<td>1152</td>
</tr>
<tr>
<td>9</td>
<td>512</td>
<td>Traditional mode</td>
<td>10</td>
<td>512</td>
<td>1.95%</td>
<td>314.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fractal mode</td>
<td>2560</td>
<td>2560</td>
<td>100.0%</td>
<td>6.14</td>
</tr>
</tbody>
</table>

The FPD fractal scan control and related fractal scan IP core studied in this paper has been directly applied to large-scale multimedia LED video display systems and OLED screen indoors and outdoors. It has the ability of gray displaying up to 12 bits (4096 gray scale), and up to 200Hz fresh frequency makes the image stable, clear and high cost-effectiveness. It convinced that the system can not only achieve high clarity and resolution of FPD video image displaying, but more important cut down the cost of FPD driving with no high-speed IC devices as well.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the National Natural Science Foundation of China for providing financial support for this work under Grant No.60773081 and Grant No.60777018, and also to acknowledge the financial support by Shanghai Science and Technology Department under Grant No.07062008 and Grant No.08706201800

REFERENCES

An Improved Technique for Object-based Video Segmentation

Wendan Xu\textsuperscript{1} and Donglai Xu\textsuperscript{2, *}

\textsuperscript{1} Department of Computer Engineering  
Xi'an Aero-technical College  
Xi'an, 710077, P. R. China

\textsuperscript{2} School of Science and Technology  
University of Teesside  
Middlesbrough, TS1 3BA, UK

ABSTRACT

This paper presents a new video segmentation scheme, which consists of two stages: initial segmentation and motion estimation. In the initial segmentation, the watershed transformation followed by a region adjacency graph guided region merging process is used to partition the first video frame into spatial homogenous regions. Then the motion of changed region is estimated. Based on the highly efficient quadratic motion model, the motion estimation is undertaken using Gauss-Newton Levenberg-Marquardt method to minimize the least-square error function. Experimental results show the proposed scheme provides high performance in terms of segmentation accuracy and video compression ratio.

1. INTRODUCTION

To support content-based functionality, the concept of the Video Object Layer (VOL) is introduced in MPEG-4 video coding standard. To adopt this concept, a video image needs to be segmented into a collection of video objects in terms of shape, color and motion information, etc. In many object-based video applications, video segmentation plays a fundamental role. Despite the fact that video objects can be easily distinguished by the human visual system, for computer to automatically and accurately segment video objects still remains one of the most challenging tasks. Hence, this problem has attracted the attentions of many researchers.

There are many techniques proposed for video object segmentation [1]. Spatial segmentation identifies spatially homogenous regions in an image scene with respect to different visual appearance, such as colors and textures. Therefore it can determine object boundaries. Temporal segmentation partitions a video frame into moving objects, since most moving objects have coherent motion characteristics that are distinctive from each other. In fact, the interdependence between motion and spatial boundary has to be taken into account in segmentation process. This is because on one hand, the object boundaries have to be known to estimate an accurate motion; on the other hand for correct object boundary detection, an accurate description of motion is necessary. Therefore, the spatial-temporal segmentation approaches seem to be the most promising solution to the problem, which effectively combine spatial and temporal segmentation methods and make use of both spatial and motion properties of video objects.

This paper describes a technique for video motion estimation and segmentation. The technique consists of two stages: initial segmentation and motion estimation. For initial segmentation, Gaussian blurring followed by a watershed transformation is used to partition the first video frame into spatial homogenous regions. Although watershed transformation delineates good region borders, it also yields over-segmentation results. To significantly reduce the over-segmentation, a region adjacency graph guided region merging process is adopted. And then the changing regions between two consecutive frames are detected, and the motion parameters are estimated for each of the changed regions. Because of its high accuracy and coding efficiency, the quadratic motion model has been chosen to do the motion estimation, rather than simplistic affine motion model.

The rest of the paper is organized as follows. Section 2 gives the description of the initial segmentation and the region adjacency graph guided region-merging method. In section 3, the details of the quadratic motion estimation method are explained. The experiments carried out and the results obtained are presented in section 4. Finally, the conclusions are drawn in section 5.

* Corresponding author: Te: +44 1642 342496; Fax: +44 1642 342401; Email: d.xu@tees.ac.uk
2. INITIAL SEGMENTATION

The segmentation scheme starts with a spatial segmentation algorithm that is applied to the first frame of a video sequence to get the initial objects. This spatial segmentation uses watershed transformation [2, 3] to obtain primitive spatial homogenous regions. Then region-merging process is used to reduce over-segmentation.

2.1 WATERSHED TRANSFORMATION

To partition the first frame of a video sequence into primitive regions through the watershed transformation, the frame is firstly smoothened using Gaussian low-pass filter to reduce image noise, then the magnitude of the image gradient is computed, which is the input of watershed transformation. Under watershed transformation, every local magnitude minima will lead to a watershed region called catchment basin through a flooding process. To obtain the local minima, the magnitude image is sorted in incremental order using an address table. In the flooding process, the pixels of the magnitude image are accessed and assigned to a label, leading to the result of labeled region image.

2.2 REGION MERGING

Since watershed transform usually produces over-segmentation, the region merging process [4, 5] is applied and it is guided by a weighted region adjacency graph. The region adjacency graph is formed at the beginning of region merging process. Let \( R = \{ r_i \} \) be a disjoint partition of input image frame, which is the initial watershed over-segmentation result, \( E = \{ (r_i, r_j) \} \) be a set of edges that links two neighboring regions \( r_i \) and \( r_j \), a region adjacency graph (RAG) can be expressed as \( G=(R, E) \). A weight is assigned to each edge in RAG. This weight is a measure of dissimilarity between the two regions linked by the same edge. It provides a merging criterion [6]. The merging criterion or dissimilarity measure is decided based on the mean luminance in this paper. If \( l_i \) and \( \mu_i \) are denoted as the area and mean luminance of region \( r_i \), respectively, and \( l_j \) and \( \mu_j \) are the area and mean luminance of region \( r_j \), respectively, then the dissimilarity measure associated with the two neighboring regions \( r_i \) and \( r_j \) is defined as:

\[
w_{ij} = \left( \frac{l_i + l_j}{l_i + l_j} \right) \cdot \left( \mu_i - \mu_j \right)^2
\]

(1)

After all of the edge weights are calculated using the above formula and the RAG is formed, the region merging process starts. The region-merging algorithm scans the RAG, searches for the edge with the minimum dissimilarity value and merges the two most similar neighbouring regions linked by an edge. This merging process is performed iteratively until a stopping criterion is satisfied. The stopping criterion can be a minimum edge-weight threshold. In each region merging iteration, in order to reflect the changes generated by merging two regions, a considerable amount of updating operations to the RAG are performed. In these operations, the step of updating edge weights consumes a lot of computation time. To speed up the algorithm, we take two measures to decrease the number of edge-weights to be updated. One measure taken here is that, when searching for the minimum edge weights, if more than one edge are found, the neighbours of two regions linked by these edges are counted. Then the two regions linked by the minimum weight edge and having the least amount of neighbours are selected to merge. Another measure is pre-absorbing the very small regions. This operation is performed before the region merging iterations, in which the very small regions whose areas are less than a predefined threshold, are absorbed by those most similar neighbours.

3. MOTION ESTIMATION

After initial segmentation is completed, the motion information of every objects needs to be estimated. Before motion estimation is performed, the changed objects are detected by the change detection processing between two consecutive frames. Based on a 12-parameter quadratic motion model [7, 8], the motion estimation process is reduced to an optimization of the least-square error function on each changed regions. This minimization problem is solved by Gauss-Newton Levenberg-Marquardt method [9].

3.1 QUADRATIC MOTION MODEL

The 12-parameter quadratic motion model [7, 8] is a set of pixel vector transforming equations, mapping a pixel coordinate within a region from one image frame to another. The model makes it possible that complex motions of a
A large moving region, such as rotation, rapid translation, and scaling, are efficiently represented by a single set of motion parameters.

The motion model can be mathematically described by the quadratic transform shown in the following equations:

\[ u(x, y) = a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x + a_5 y + a_6 \]

\[ v(x, y) = b_1 x^2 + b_2 y^2 + b_3 xy + b_4 x + b_5 y + b_6 \]

where \((x, y)\) is the coordinate of an image pixel in image frame \(t\); \((u(x, y), v(x, y))\) is the corresponding image coordinate in image frame \(t+1\); and \((a_1, \ldots, a_6, b_1, \ldots, b_6)\) are the quadratic transform parameters.

### 3.2 Motion Estimation

If \(W(x; p)\) denotes the parameterized motion transformation, the goal of motion estimation is to minimize the following least-square error function between two image frames:

\[
E(p, R) = \sum_{x \in R} \left[ I(W(x; p)) - 1(x) \right]^2
\]

where \(p = (p_1, \ldots, p_n)^T\) is a motion parameter vector; \(R\) is the region where the \(p\) is calculated; \(x = (x, y)^T\) is the coordinate vector of a pixel in the region \(R\); and \(I\) represents image intensity.

To minimize the expression in Eq. (2), an initially estimated \(p\) is assumed and given. Then the following expression is minimized to iteratively solve the increments to the parameters \(\Delta p\) [10, 11]:

\[
E(\Delta p, R) = \sum_{x \in R} \left[ I(W(x; p + \Delta p)) - 1(x) \right]^2
\]

and the parameter \(p\) can be updated as follows:

\[
p \leftarrow p + \Delta p
\]

The expressions (3) and (4) are iteratively computed until the estimation of the motion parameter \(p\) converges. The convergence condition can be that, a norm of the vector \(\Delta p\) is below a given threshold \(\varepsilon\), i.e. \(\|\Delta p\| < \varepsilon\), or simply the increment of translation parameters are less than a predefined threshold, or even the iterative count is up to a designed limit.

To optimize function (3), the Gauss-Newton Levenberg-Marquardt method is used. A first order Taylor expansion on \(I(W(x; p + \Delta p))\) is performed initially as follows:

\[
1(W(x; p + \Delta p)) = 1(W(x; p)) + \nabla I \frac{\partial W}{\partial p} \Delta p
\]

where \(\nabla I\) is the gradient of image \(I\), which is evaluated at \(W(x; p)\), and \(\partial W / \partial p\) is Jacobean matrix of the motion transform.

To obtain the optimized \(\Delta p\), the partial derivative of the expression in Eq. (3) with respect to \(\Delta p\) is set to zero:

\[
g = \frac{\partial E(\Delta p, R)}{\partial \Delta p} = 2 \sum_{x \in R} [\nabla I \frac{\partial W}{\partial p}]^T \left[ I(W(x; p)) + \nabla I \frac{\partial W}{\partial p} \Delta p - 1(x) \right] = 0
\]

If we solve Eq. (6), the closed minimum of \(\Delta p\) is as follows:
\[
\Delta p = -H_{\text{LM}}^{-1} \sum_{x \in a} \left[ \nabla I \left( \frac{\partial W}{\partial p} \right)^T [t(W(x; p)) - I(x)] \right] \tag{7}
\]

where HLM is a Levenberg-Marquardt approximation to the Hessian matrix:

\[
H_{\text{LM}} = \sum_{x \in a} \left[ \nabla I \left( \frac{\partial W}{\partial p} \right)^T \left( \frac{\partial W}{\partial p} \right) + \delta \sum_{x \in a} \text{Diag} \left( \left( \nabla I \left( \frac{\partial W}{\partial p} \right) \right)^2 \right) \right] \tag{8}
\]

where \text{Diag}(\ ) is the Levenberg-Marquardt diagonal matrix; \(\delta\) is the parameter to adjust the approximation of Hessian matrix. Starting with a small value 0.01, \(\delta\) is updated upon the provision of evaluated error after each iteration. The updating rule is that, if the error decreases, \(\delta\) is reduced by ten times and \(\Delta p\) is updated; otherwise \(\delta\) is increased by ten times in next iteration.

4. EXPERIMENTAL RESULTS

To analyze the performance of the proposed segmentation scheme, experiments have been carried out based on several MPEG-4 standard video sequences (QCIF format, i.e. 176 pixels × 144 lines). Figure 1 shows the 95th and 100th frames of MPEG-4 video sequence “Claire”. Using the changed detection algorithm, the change of the head between the two frames is recognized, as shown on the left in Figure 2. On the right of the Figure 2 is the region merging result. Here we assign the stop criterion of the region merging process being the six regions obtained, and all the very small regions whose area are less than 20 pixels are pre-absorbed before the merging process. And then the motion estimation is undertaken and the warped results, which correspond to the 100th frame moved from the 95th frame, are acquired. In Figure 3, the left image is the changed head part between the 95th and 100th frame; in the middle is the warped head from the 95th frame based on the estimated affine motion parameters, on the right is the warped frame from the 95th frame. Figure 4 shows the result based on quadratic motion estimation. From the comparison between Figure 3 and Figure 4, we can see that the motion estimation based on quadratic motion model improves accuracy over the affine motion estimation. In fact, the PSNR values between the warped frames and the 100th frame are also computed here. For the affine model and quadratic motion model, PSNR is 42.5db and 44.6db, respectively.
5. CONCLUSIONS

An improved video segmentation scheme has been proposed in this paper. The scheme consists of two stages: initial segmentation and motion estimation. The initial segmentation starts with a watershed transformation on the first frame of a video sequence. Because the watershed transformation produces the over-segmentation result, a region adjacency graph guided region merging process is then used to obtain the initial segmentation objects. In the region merging, two measures are taken to speed up the algorithm. In the motion estimation stage, the motion parameters of changed region are calculated. Based on the highly efficient quadratic motion model, the motion estimation process is simplified to be the minimization of a least-square error function, which uses Gauss-Newton Levenberg-Marquardt method. Experimental results demonstrate that compared to the widely used affine motion model, the proposed scheme provides high accuracy in terms of region segmentation, and leads to high coding efficiency.

REFERENCES


Critical Infrastructure Protection Security Layer for DNP3 Devices

Todd Mander¹, Farhad Nabhani² and Ivan Lee³

¹Ryerson University
Toronto, Canada

²University of Teesside
Middlesbrough, U.K.

³University of South Australia
Mawson Lakes, Australia

ABSTRACT

A security layer for the popular utility Distributed Network Protocol (DNP) is proposed, limiting the effectiveness of cyber attacks on critical infrastructures with essential confidentiality. Ongoing automation of critical infrastructures, including power, gas, and water, increase the responsiveness of control operations and increase the number of utility devices. This ongoing automation therefore increases cyber attacker capabilities to disrupt utility operations and therefore create serious security challenges. These security challenges result in manufacturer challenges that are not readily met by current commercial security implementations. In addition, commercial security implementations are not designed for utility characteristics. The security layer presented in this paper meets the security challenges, manufacturer challenges, and handles utility characteristics. The security layer supports multiple security ciphers per connection, eliminating cipher re-negotiations that would delay time-critical data. Encryption ciphers are combined into recipes to handle decreased effectiveness of ciphers over the long lifespan of devices. Multiple security levels allow immediate transition to stronger security during attacks and transition to less intensive security during normal operations without negotiations. In addition, the security layer provides quality of service selectors to tailor the security requirements of data transmissions. This provides differential security for data classified as high value to attackers without requiring all data utilize the same level of security.

1. INTRODUCTION

The security layer presented in this paper for critical infrastructure security is a sublayer of the DNP3 application layer and provides full confidentiality for DNP3 application layer fragments. The security layer does not provide authentication, which is provided by the DNP3 secure authentication standard [1]. Critical infrastructures, including electricity, gas, and water, increasingly require security as the number of devices that can be attacked increases as well as the number of reasons for attacking devices increase [2]-[8]. Attacks on the utilities can result from consumers attempting to defraud the utilities, to organized crime attempting to influence revenue or contractual obligations, to terrorists attempting to create sustained service disruptions. DNP3 is a popular utility protocol used worldwide, including by the Water Industry Telemetry Standards (WITS) group in the United Kingdom [9]. However, DNP3 currently only specifies authentication security [1]. This allows attackers to, at minimum, gather intelligence on device attributes, operating states, system weaknesses, and statistics on communication failures for denial of service attacks. To counter-act these threats more effectively confidentiality is required.

There are many commercial security solutions available to utilities, such as Transport Layer Security (TLS) and Internet Security (IPsec) [10]-[11]. However, these commercial implementations are not designed for the critical infrastructure characteristics or meet the manufacturing challenges that the characteristics create. Three main manufacturing challenges for utility devices include: very large numbers of devices, long equipment lifespan, and constrained processing capabilities [2]-[3], [8], [12]. Utilities may operate in excess of millions of devices to monitor the system, such as smart meters [12]. Attempts to update or replace large numbers of security routines within short periods of time will therefore create large bandwidth requirements not typically required by the utilities. Typical utility data transmission intervals are between 2 seconds or less for the transmission system and 60 minutes for smart meters [6], [13]-[14]. As a result, devices using a single security routine for a connection, such as with TLS or IPsec, may unnecessarily increase bandwidth requirements. The security layer utilizes multiple security routines, thereby ensuring that replacing security routines will not interfere with typical utility traffic or require higher bandwidth to support the security layer operations. Utility devices are typically designed for proven technological platforms. Therefore, devices manufactured for utilities are unlikely to contain leading edge processors or contain the latest features expected in personal computers. As a consequence, the security layer must provide flexibility to reduce the processing overhead when possible. Flexibility is achieved through the security levels and the quality of service (QoS) selectors that ensure only the necessary security is applied to the data rather than a general application of stronger and more process intensive security for all data transmissions.
2. Security Layer Databases and Parameters

2.1. Security Routines

The security layer encryption operations applied to the DNP3 application layer data fragments utilize security routines similar to Transport Layer Security (TLS) records or Internet Security (IPsec) security associations (SAs) [10]-[11]. The security routines differ from TLS and IPsec implementations by utilizing multiple security routines associated with specific selectors rather than a single security routine for a specific set of selectors.

The security layer is cipher independent, allowing changes in supported cipher algorithms without alterations to the security specification, which mirrors IPsec and TLS development [10]-[11]. The security layer is capable of supporting a wider selection of encryption algorithms than IPsec and TLS, including asymmetric, symmetric, block, and stream ciphers. The security layer does not support authentication since authentication is implemented by the DNP3 secure authentication specification [1]. The security layer encryption covers the entire DNP3 application layer fragment, shown in Figure 1, and prefixes the code number (CN) onto the encrypted application layer fragment. The code number provides a reference to the destination device indicating which security routine was used on the application layer fragment. The code number is equivalent to the IPsec security parameters index (SPI) except that it is bidirectional [11].

![Figure 1: Encryption of a DNP3 application layer fragment.](image1)

The code number is referenced from the security parameter database (SPD) that provides the necessary parameters for the encryption operations to encrypt or decrypt the application layer fragment. The SPD contains entries indicating the cipher type, key length, cipher mode (block ciphers), private key (asymmetric and symmetric encryption), public key (asymmetric encryption), expiration time, and general parameters such as initialization vectors. There are also other SPD parameters presented in the other sections of this paper.

An advantage of the multiple security routines is that there is less data available to attackers for attempting to break the encryption. However, utility devices may be constrained in the number of security routines that can be supported to significantly increase security, e.g. due to memory limitations. The effective number of security routines can be increased with the use of aliases to counter device limitations. A security routine can have multiple aliases, increasing the apparent number of code numbers from the attacker’s viewpoint with minimal overhead. The alias database (AD) is an array of one or more alias code numbers associated with a particular SPD code number.

2.2. Security Routine Recipes

A major contrast between devices manufactured for utilities and general commercial devices is the device lifespan, which may be 15 to 20 years for the utilities [8]. As a consequence of the device lifespan, the integrity of the ciphers cannot be guaranteed. Current industry accepted ciphers may become obsolete within the device lifespan due to new devised attacks. Cipher algorithm replacement may be difficult due to the large number of deployed devices, and if the algorithms are FPGA or firmware based. The security layer increases the effectiveness of the weakened ciphers by implementing a recipe format for the security routines shown in Figure 2.

![Figure 2: SRRD security routine recipe encryption of a DNP3 application layer fragment using SRD entries.](image2)
The recipe format allows weaker security routines to be used consecutively on the same application layer fragment, increasing the overall effectiveness of the security to current accepted levels. Alternatively, the security routine recipes can provide stronger security during attacks without requiring additional stronger security routines to be stored on the device. The security routine recipe database (SRRD) provides a unique code number containing a list of security routines from the SPD. The encryption operations follow the list from left to right and the decryption operations follow the list from right to left as shown in Figure 2. An advantage of the security routine recipe is that the recipe does not have to contain the same cipher types, cipher keys, or cipher modes. This greatly increases the number of security options for the application layer fragments, thereby increasing the difficulty level for attackers to determine the structure of the security operations and break the encryption. The security routine recipes also utilize aliases equivalent to the security routines, which are also contained within the alias database (AD).

3. SECURITY LAYER OPERATIONS

3.1. SECURITY MODES

The multiple security routines and recipes provide quality of service (QoS) capabilities to DNP3 devices, increasing the responsiveness of the security layer to current network security requirements. As a result of the multiple security routines, utility devices can implement the least process intensive security for typical traffic without compromising the overall security of the network. When the network experiences an attack, the devices can immediately transition to stronger security without further negotiation, as would be required for TLS or IPsec, due to the multiple security routines and recipes. The security layer groups the security routines and recipes into security levels that provide the appropriate security response to current network conditions. The security layer implements five security modes for handling the security levels, which are shown in the state machines in Figure 3 and Figure 4. The five states are: None, Low, Normal, High, and Extreme.

The first received data transmission by the device will cause the security layer to transition from the initial None state to a security mode that matches the security level of the received data transmission. The Low state, or low security level, is used for minimal security requirements beyond the DNP3 authentication security. The Low state contains weaker security, and is used to only prevent non-intensive security attacks until the protected data becomes obsolete. The Low state may also be used for low risk environments, such as devices communicating within the same facility. Any security mode above the Low state may use security routine recipes. The Normal state, or normal security level, is the typical security state for communication between the source and destination, and uses typical industry accepted ciphers. The High state, or high security level, increases security to counter-act current threats within the network or to protect high value data. The high security level utilizes stronger ciphers and cipher keys than may typically be used, such as using 256-bit instead of 128-bit encryption. The high security level security routine recipes typically consist of Normal state security routines. The Extreme state, or extreme security level, significantly increases security during sustained attacks on the network. The Extreme state utilizes very
strong security, such as a cipher’s maximum key length and security recipes containing High state security routines. In the situation where a significant amount of Normal state security routines are retired unexpectedly, the High and Extreme states become trap states until the Normal state security routines are replaced.

The security mode operations require an additional entry in the security parameter database (SPD) and the security routine recipe database (SRRD). The additional entry indicates which security routines and recipes are associated with each security mode. The security routines and recipes may be associated with more than one security mode, e.g. a manufacturer may consider AES-256 encryption to be typical security and sufficient for the higher security levels, thereby avoiding increased processing overhead for more intensive ciphers.

Two similar state machines are shown in Figure 3 and Figure 4. Figure 3 shows the state machine for lower nodes in the network hierarchy communicating with higher nodes, such as data concentrators communicating with control centers. This is called the outstation side (OS) state machine. Figure 4 shows the state machine for higher nodes in the network hierarchy communicating with lower nodes, such as data concentrators communicating with outstations. This is called the master side (MS) state machine. The implementations of the state machines are nearly equivalent, with the MS state machine typically following the OS state machine transitions.

![Figure 4: Security mode master side (MS) state machine including interactions with OS state machine.](image)

There are two transition types within a security mode state machine: external stimulus and internal stimulus. The external stimulus transitions are security layer responses to received code numbers from encrypted application layer fragments. If the received code number’s corresponding security mode is equivalent to the current security mode, there is no transition between states for both Figure 3 and Figure 4. However, if there is a change in the corresponding security mode to the received code number, the OS and MS state machines transition immediately to the new matching higher security level. The internal stimulus parameters will determine if the OS and MS state machines will transition down to a lower security level if the new corresponding security mode is lower.

The security layer external stimulus transitions eliminate additional security protocol communication overhead that would typically be required to manage the security levels of devices, and thereby provides critical support for utility devices. The OS state machine in Figure 3 provides a ripple effect security response (RESR) from the control center in communicating the new security levels to the network. The control center transmits an encrypted frame corresponding to the new security level to the devices it normally directly communicates with. These devices transition to the corresponding security mode, which causes a similar transition in the MS state machine. This MS transition passes on the corresponding security level to the devices it communicates with in a ripple effect until the entire network is at the desired security level set by the control center. This security level update method ensures that the control center minimizes the number of devices it communicates with, instead of being required to communicate with every single device in the network. The RESR thereby reduces the complexity of the security layer and reduces the bandwidth requirements.

The internal stimulus transitions are security layer responses to local device conditions, such as detected tampering or DNP3 authentication challenge failures [1]. The internal stimulus transitions will cause the state machines to transition to one security mode higher than the current security mode. For the OS state machine, the higher security level will be maintained by the internal stimulus while the threat persists regardless of the external
stimulus (received code numbers). However, the OS state machine will drop to a lower security mode if an external stimulus caused the OS state machine to transition to an even higher security level than required by the internal stimulus threat. As a result of eliminating the security layer protocol overhead for communicating security modes to devices, the MS state machine requires additional overhead in comparison to the OS state machine. The MS state machine must occasionally retest the external stimulus from an outstation (the outstation’s internal stimulus) to ensure that the higher security is still required. Otherwise the outstation’s OS state machine would continue to follow the master’s higher security whether it was needed or not. The MS state machine will attempt to use a lower security level following its next poll or response to the outstation, which is shown in Figure 5. The OS state machine may employ multiple security mode transitions if there is a compound internal stimulus threat, e.g. unacceptable authentication challenge failure rates to cause a transition from Normal to High to Extreme states.

Figure 5: MS state machine response timeline to an outstation internal stimulus threat response.

Once the None state is left after security level initialization for both the OS and MS state machines, it cannot be re-entered until the security layer is re-initialized with replacement security routines and recipes. The Normal state cannot be re-entered from the higher security mode states if the Trap flag has been set, indicating that the security modes have been trapped in the High and Extreme states until the security layer is re-initialized. The trap flag ensures that the security layer maintains sufficiently strong security when supported ciphers are weakened or when a significant number of lower security cipher keys have expired.

3.2. Security Layer Selectors

The security layer implements quality of service (QoS) selectors similar in concept to IPsec selectors [11]. The QoS selectors provide additional control over the level of security used on an application layer fragment. The security layer supports three types of selectors: function code, data object, and user. These three QoS selectors may temporarily override the security mode security routine and recipe selection but does not affect the security mode operations.

DNP3 function codes [15] are a coarse QoS selector to determine which application layer fragments require special security considerations, either due to their associated data’s high value to attackers or their lack of value to attackers. Function code QoS selectors are very device specific, except for DNP3 authentication security which generally requires stronger than typical security. A flag value is therefore used to indicate if a higher security level is required than the current security level when authentication function codes and objects are used. The function code QoS selector database (FCSD) contains the parameters for applying the function code QoS selector. The FSCD additionally contains user identifications for handling the user QoS selector. The FSCD contains function code value ranges to reduce the size of the FSCD, by eliminating the need to potentially create an entry for each function code. A security level entry indicates whether the current, next lower, or next highest security mode should be used. The FSCD allows the security layer to specify the security types allowed for the function code, e.g. if only standard government ciphers are allowed. An entry also indicates if asymmetric ciphers are allowed to be used with the function code QoS selectors.

DNP3 data objects [16] are a finer QoS selector, and provide additional security for data objects that are of particular value to cyber attackers, such as device self-descriptions. It is not practicable to define many data object QoS selectors due to the large number of DNP3 defined data objects [16]. Therefore, only critical data objects can be defined as selectors. The data object QoS selector database (DOSD) is much simpler than the FCSD and contains the list of data objects requiring stronger encryption and the allowed ciphers for that data object similar to the FSCD.

The user QoS selector provides support for the DNP3 authentication security user identifications [1]. This QoS selector provides special security considerations based on user requirements, e.g. if the user accesses data of low value to attackers then less process intensive security can be utilized. The user QoS selector can either use the
general mode or the function code mode. The general mode uses security routines and recipes that are specific to the user only. This requires an additional entry in the SPD and SRRD to indicate if the security routine or recipe is associated with a particular user. The function code mode is used for the user selector when only specific function codes require special consideration for the DNP3 user identification. The function code mode uses the FCSD database with the user identification as the reference.

3.3. SECURITY LAYER SECURITY ROUTINE SELECTION

The security layer implements multiple options for selecting a security routine or recipe for application layer fragment encryption based on the security modes and the quality of service (QoS) selectors. The selection procedure to determine the appropriate security routine to be used for the encryption is shown in Figure 6.

In Figure 6, the security layer selection process is based on the source/destination address pair and collects the required parameters to finalize the selection of the security routines and recipes. If the security mode is in the Extreme state, only the security routines corresponding to the Extreme state are selected and sent to the code number selection process.

If the security mode is not in the Extreme state, the security layer uses the DNP3 user identification. If the user selector general mode is used, the user identification value is passed to the security mode selection process. If the function code mode is used, the user identification value is passed to the function code selection process. If the user identification is not defined, no values are passed to the function code selection process. The function code selection process determines if the function code is listed in the FSCD, or if the authentication flag is set. If the function code is located or if the flag is set, the requested security level is passed to the security mode selection process and the list of allowed ciphers if defined. Otherwise the data object selection process is entered. If the application layer fragment contains data objects that are defined in the DOSD, the data object selection process passes the requested security level to the security mode selection process and the list of allowed ciphers if defined. If there are no defined data objects, no parameters are passed to the security mode selection process.

The security selection mode process will select all possible security routine and routine recipes for the security mode from the SPD and SRRD. If there are no received parameters, the current security mode will be used. Otherwise the received parameters will alter the security mode for the selection process to the current, one lower, or one higher security mode. If the parameters include the user identification, only the security routines and recipes for that user will be selected from the security mode. If there is a list of allowed ciphers, the list of selected security routines and recipes will be reduced to contain only those ciphers. The resultant list of possible code numbers are sent to the code number selection process. One of the code numbers will be selected at random and used for the application layer fragment encryption. The selected code number value is sent to the alias selection process, which determines the list of possible alias code numbers for the selected code number. If aliases are available, one of these values will be selected at random for the code number to appear within the application layer fragment.

3.4. SECURITY LAYER MAINTENANCE

Similar to IPsec, the method for updating security routines and recipes are outside the scope of the security layer. This allows manufacturers to implement maintenance protocols specific to device requirements, such as automatic updates based on Internet Key Exchange (IKE) [17] or manual updates. The code number in the application layer fragment can only represent 256 security routines, security routine recipes, and aliases if one byte is used for the code number. Using more bytes creates a very sparse numbering system in which pattern detection regarding current security routines, security routine recipes, and aliases may become easier for attackers. Therefore, the
security routines, security routine recipes, and aliases are used in blocks, where one block will replace the current block after expiration. The security blocks overlap to provide a transition phase, reducing the possibility of state errors, e.g. due to delayed transmissions using an older code number. The transition phase also deemphasizes the transition period so that attackers are unable to clearly delimit when a code number represents a new security routine or recipe. This decreases an attacker’s capability to effectively process and break ciphers.

The protocol for handling the security routine and recipe activation and deactivation is shown in the state machine in Figure 7. Initially a security routine or recipe is valid but inactive. Once the security routine or recipe is initiated, as determined by the SPD and SRRD, the security routine or recipe becomes active and can be used by the security layer. After the security routine or recipe expiration occurs, it becomes invalid and will not be used for transmitted data. However, the security routine or recipe may be used for received data due to potential unexpected network delays. This grace period prevents unnecessary time-critical data retransmissions and eliminates security layer protocol overhead in ensuring that encrypted data arrives before code number expiration. After a set length of time, depending on the source/destination transmission latency, the invalid security routine or recipe is deleted.

Figure 7: Security routine and recipe life cycle.

4. SECURITY LAYER SAMPLE CONFIGURATION

Utilities have wide varying security requirements for devices, and therefore require the flexibility implemented by the security layer. An example comparing an electrical utility substation device with a smart meter is presented in this section. The substation device will be within a high traffic environment and transmitting approximately every 2 seconds. The smart meter may be transmitting once every 60 minutes and is therefore in a low traffic environment. The substation device will be readily accessible by personnel unlike the smart meter. Based on these basic parameters, a sample security layer configuration is presented.

Table 1 provides the basic configuration information for the security layer maintenance parameters. The replacement cycle period, in which all security routines and recipes are updated for the device, is relatively quick for the substation due to the high traffic volume. The replacement rate for the smart meter is much lower due to the very low typical traffic volume. The substation security layer can be updated manually since there are personnel available to perform the updates. However, automatic updates can be employed under various circumstances, including providing security for temporary device access by personnel. The smart meter is updated manually due to the low traffic volumes and to ensure a yearly maintenance check, including detecting tampering by customers. The block cycle period is less than a week for the substation due to the high traffic volume and the relative ease to update the security routines and recipes. The smart meter block cycle period is longer due to the low traffic volume and to decrease the number of security routines and recipes that must be stored in the smart meter. The block transition time must be short in order to decrease the number of security routines and recipes contained in higher speed memory, but long enough to ensure that block transitions cannot be easily determined by attackers. The number of blocks used is based on the replacement cycle period divided by the block cycle period plus one. The extra block provides contingency against delayed maintenance updates.

Table 1: Maintenance Parameter Configuration Profiles

<table>
<thead>
<tr>
<th></th>
<th>Substation</th>
<th>Smart Meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replacement Cycle Period</td>
<td>&lt; 4 weeks</td>
<td>Yearly</td>
</tr>
<tr>
<td>Block Cycle Period</td>
<td>&lt; 1 week</td>
<td>1 month or greater</td>
</tr>
<tr>
<td>Block Transition Time</td>
<td>24 to 48 hours</td>
<td>1 to 4 days</td>
</tr>
<tr>
<td>Number of Blocks in Cycle Period</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 2 provides the basic configuration information for the security modes. The substation does not support the process intensive asymmetric encryption due to high traffic volumes. The smart meter supports asymmetric encryption for the QoS selectors discussed in Table 3. Substation code number alias support is limited in order to support more security routines and recipes since the substation is a high value target for attackers. The smart meter provides increased support for aliases to limit the number of stored security routines and recipes. For the security
mode Low state, the substation device would use standard ciphers but in the less secure but more efficient electronic codebook mode for the block ciphers. The smart meter would use less secure ciphers to only delay attackers. The Normal state would use standard ciphers, such as in the 128-bit key range, for both the substation and smart meter. However, the substation would also contain more secure ciphers such as 256-bit key ciphers. The High state would use more secure standard ciphers, such as in the 256-bit key range for both the substation and the smart meter. However, the smart meter will be more reliant on security routine recipes consisting of Normal state security routines. The Extreme state for the substation would use the maximum allowed keys sizes for the ciphers but mainly use security routine recipes consisting of the High state security routines. The smart meter for the Extreme state would use more secure ciphers, such as those in the 256-bit key range.

Table 2: Security Mode Configuration Profiles

<table>
<thead>
<tr>
<th></th>
<th>Substation</th>
<th>Smart Meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetric Encryption</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Aliases</td>
<td>Minimal amount</td>
<td>Significant amounts</td>
</tr>
<tr>
<td>Low State</td>
<td>128-bit ciphers, ECB mode</td>
<td>Older ciphers, DES</td>
</tr>
<tr>
<td>Normal State</td>
<td>Mostly 128-bit ciphers some 256-bit</td>
<td>128-bit ciphers</td>
</tr>
<tr>
<td>High State</td>
<td>Mainly 256-bit ciphers, some security recipes containing Normal state ciphers</td>
<td>Some 256-bit ciphers, mainly security recipes containing Normal state ciphers</td>
</tr>
<tr>
<td>Extreme State</td>
<td>Some high bit ciphers, mainly security recipes containing High state ciphers and more complex Normal state ciphers</td>
<td>Mainly 256-bit ciphers, some more complex security recipes containing Normal state ciphers</td>
</tr>
</tbody>
</table>

Table 3 provides the basic configuration information for the QoS selectors. The substation device does not provide special consideration for application layer fragments since relatively high security is already implemented. As a result, additional security for DNP3 authentication security and function codes is not required. The smart meter does not provide as much security as the substation device; therefore higher security is needed for data transmissions that utilize the DNP3 authentication security. The smart meter utilizes asymmetric encryption for read/write function codes since this would typically represent revenue services, e.g. accessing the KWh counters. The asymmetric encryption provides non-repudiation for the encrypted data so that claims regarding electrical consumption are more strongly addressed. This function code QoS selector uses the current security mode but with only asymmetric ciphers. The device attribute data objects [16] will be of particular value to attackers and therefore requires additional security for both devices. To maximize the number of available security routines and recipes for the substation device, user QoS selector security must be minimized. The user QoS selector is not used for the smart meter to limit the number of security routines and recipes that must be supported.

Table 3: QoS Selector Configuration Profiles

<table>
<thead>
<tr>
<th></th>
<th>Substation</th>
<th>Smart Meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNP3 Authentication</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Function Code Security</td>
<td>No</td>
<td>Read/Write (Asymmetric ciphers)</td>
</tr>
<tr>
<td>Data Object Security</td>
<td>Device attributes</td>
<td>Device attributes</td>
</tr>
<tr>
<td>User Security</td>
<td>Limited</td>
<td>No</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The security layer presented in this paper provides very flexible security to handle various device requirements within critical infrastructure networks. The security layer provides full confidentiality for DNP3 application layer fragments on an end-to-end basis to complement the DNP3 authentication security [1]. However, the security layer does not provide security coverage for the DNP3 transport function and data-link layer. Although these layers are not as critical to DNP3’s operations, they can be used in denial of service attacks. To counteract this threat, the DNP3 device communication links should be isolated from attackers or additional security below the data link layer employed on vulnerable links, e.g. with IPsec.

The multiple security routines and recipes meet manufacturing challenges for the utilities. Large volume cipher updates within a small period of time is not practicable due to the large number of devices and the typical limited bandwidths used for utilities. The security routines and recipes solve this problem by ensuring that a longer transition period is available for updating security. A cipher lifespan cannot be guaranteed within a device’s lifespan and therefore may become eventually ineffective. Replacing the cipher algorithm may not be practicable.
The security routine recipes increase the effective lifespan of the cipher algorithms by combining security routines with various cipher types and key lengths, increasing the effective security of weaker ciphers. The security modes ensure that there is immediate stronger security available for data during attacks. As a result, critical infrastructure systems are not risked due to cipher renegotiations that would delay time-critical data. In addition, lower security levels ensure that normal traffic is not delayed by more process intensive security in the higher security levels.

The QoS selectors provide additional flexibility to the security layer, either to increase or decrease the security for specific data. Alternatively, the QoS selectors provide special security considerations, such as the asymmetric cipher non-repudiation capabilities in the smart meter example. The QoS selectors must be carefully employed to minimize the number of selectors used since they can create large databases that may not be capable of being supported by utility devices with limited processing capabilities. It is also possible for the security layer QoS selectors to actually increase the number of potential code numbers and hence the amount of security layer processing, e.g. if a lower security mode is transitioned to a higher security mode that contains more security routines and recipes than the lower security mode. Although this type of constraint is not applicable to the IPsec selectors, due to fewer possible selectors [11], IPsec selectors are not designed for utility characteristics handled by the security layer in this paper.

6. SUMMARY

The security layer presented in this paper provides very flexible security to handle disparate device requirements within critical infrastructure networks. Security has become critical due to the increased number of attackers and devices that can be attacked, from consumers attempting to commit fraud to terrorists creating sustained infrastructure disruptions. The security layer provides full confidentiality for DNP3 application layer fragments plus QoS for specific security considerations. The QoS extends the security capabilities so that specific data can utilize increased security without causing the device to unnecessarily employ increased security for typical data transmissions. The security layer complements the DNP3 authentication security and is a component for creating complete security systems that may include data security and tampering detection.

REFERENCES

Role Based Access Control Infrastructure for Consumer Demand Side Management Internet Access

Todd Mander¹, Farhad Nabhani² and Ivan Lee³

¹Ryerson University
Toronto, Canada

²University of Teesside
Middlesbrough, U.K.

³University of South Australia
Mawson Lakes, Australia

ABSTRACT

This paper proposes a new Role Based Access Control (RBAC) infrastructure to surmount the challenges resulting from the increased need to implement aggressive demand side management (DSM) for residential consumers due to declining fossil fuel based generation capacity. DSM has migrated beyond basic conservation to actively engaging consumers, e.g. allowing the consumer to audit power consumption remotely. Advances in DSM now allow utilities to remotely access and control residential devices such as air conditioning. Dispersed generation provides new avenues for the utilities to implement load shaping by remotely controlling the generation impact on the distribution system. These DSM advances critically require security to ensure that the consumer’s rights, within the home and remote access to the home, are balanced with the utility’s requirements for operating the power system. This balance can be achieved through a RBAC infrastructure. Unlike other RBAC implementations, the RBAC security proposed in this paper is designed specifically for power system characteristics. Fewer users and roles are supported in a two-tier infrastructure that differentiates roles among users by parameter sets rather than creating additional roles for a system that may contain potentially millions of devices. As a result of the fewer roles, resource conflict is handled through RBAC priority, locality, concurrency, and temporal contexts. Priority determines the preferential treatment for data access and provides the capability to block resources. Locality differentiates between local and remote users. Concurrency ensures operation accordance among users, and temporal context blocks role usage at specific times or for a specific duration.

1. INTRODUCTION

Role based access control (RBAC) security is proposed in this paper to surmount the critical security challenges for electric utilities resulting from increased stakeholder involvement with the utility devices due to more aggressive demand side management (DSM) [1]-[3]. Aggressive DSM has resulted from declining fuel reserves and the desire to avoid costs associated with additional nuclear generation [4]. For DSM to succeed, consumers must be actively engaged to increase their DSM participation and awareness. DSM is migrating beyond simply using smart metering for increasing awareness to utilities actively controlling devices within residences, such as turning on and off air conditioners and swimming pool heaters. These DSM programs are the basis for more aggressive DSM, including consumer remote monitoring of their home to utilities utilizing programmable controls for load shaping at peak demand, rather than simpler turning off air conditioning. Residential dispersed generation is adding new dimensions to DSM. Consumers can be enticed to reduce their consumption, thereby increasing their generation, to create more personal revenue. Maximizing the benefits of DSM requires that residential consumers have the capability to remotely access their homes for near real-time monitoring and control operations.

DSM remote access to residential locations creates immense security challenges and data access rights that can be resolved through RBAC security. The RBAC infrastructure presented in this paper is therefore developed specifically for electricity power system demands, and in particular for the local distribution system, handling the power system characteristics that are not handled by typical RBAC models. With remote access, the utility operations may conflict with the homeowner operations for supervisory control and data acquisition (SCADA) devices, e.g. the utility turning off air conditioning when the homeowner wants it on. Operations from multiple homeowners may conflict with each other. Additionally, the remote access by the utilities and homeowners may conflict with the local access of people currently within the residence. The RBAC developed in this paper resolves the conflict among the multiple stakeholders who have varying degrees of control and access rights to the DSM operations. The RBAC infrastructure in this paper utilizes common RBAC methods in addition to features developed for power system DSM, such as priority, locality, concurrency, and temporal contexts.
2. SCADA ORIENTED RBAC INFRASTRUCTURE

2.1 TWO-TIER MODEL

RBAC requirements for SCADA networks differ from typical RBAC models, and for which typical RBAC models would create additional SCADA device manufacturing challenges. SCADA networks potentially consist of millions of devices [1], [5], many of which are developed for low-speed access. As a result, there is limited support or capability for high user traffic volumes, e.g. typical smart metering functions may only be accessed once an hour [6]. Additionally, SCADA devices have limited processing capability, limiting the support capability for large amounts of roles and users. It is therefore vital to minimize the SCADA device access by decoupling the corporate network RBAC from the SCADA network RBAC, creating a two-tier RBAC model. Corporate network RBAC roles are condensed into fewer roles that access the SCADA network, e.g. similar in concept as a group role [7]. These condensed roles become virtual users to the SCADA network, such as DNP3 authentication security users [8], and are assigned to SCADA RBAC roles as shown in Figure 1.

![Figure 1: Two-Tier SCADA RBAC Infrastructure with Corporate and Direct Access to SCADA Roles](image)

In Figure 1, the virtual user reduces the amount of users physically accessing the SCADA network. Direct access users are corporate users or third-party users via the SCADA network or Internet, e.g. consumers accessing DSM operations. The corporate RBAC is typical RBAC security while the SCADA RBAC implements the RBAC security discussed in this paper.

2.2 RBAC INFRASTRUCTURE MODEL

Utilities typically have very few users that are allowed access to the SCADA network, which is further limited by the two-tier RBAC model. However with aggressive DSM, there may be potentially millions of users accessing the SCADA devices, e.g. residential consumers. Typical RBAC is defined for many users accessing the exact same data or functions through the exact same roles with context constraints [9]-[16]. SCADA network requirements differ from typical RBAC since there are many users accessing similar data in nearly the exact same manner. Therefore this paper uses a RBAC infrastructure to highlight the differences with typical RBAC models.

A SCADA role executes the exact same data object access throughout the network except that instead of accessing the exact same data in the exact same manner, similar data is accessed in nearly the exact same manner. The SCADA roles are differentiated by parameter sets that precisely define a role’s capabilities, e.g. limiting a consumer to their own DSM operations. The RBAC infrastructure defines three role types: universal, global, and local. The universal role type is applicable to every SCADA device with the exact same parameters, and is assigned to the utility system operators. The global role type is used for multiple device access with the exact same or nearly the same parameters, and is assigned to electricity retailers. The local role type is defined for a specific device only, and is assigned to consumers. This infrastructure reduces manufacturing challenges since manufacturers only need to define basic roles, with the specific parameters being determined by the utility.

3. RBAC INFRASTRUCTURE CONTEXT PROPERTIES

3.1 PRIORITY

The limited amount of users and roles will create data object access and control conflicts, e.g. pre-empting another user’s control. Therefore, the RBAC infrastructure resolves conflicts using priority context operations with
a basis similar in concept as [13]. The priority context interaction among the users and roles determines if the data object can be blocked from other users, blocked for a limited duration from other users, or if the data object cannot be blocked from other users. Additionally, the priority context determines preferential access to data objects when simultaneous access requests are made.

The priority access and control of a data object is a function of both the user and the role that the user is using. The RBAC priority context assigns a priority level to each role, which are: low, normal, and high. Each user is assigned a priority weighting factor for each priority level to indicate the user’s relative importance to the role, i.e. differentiating the user’s priority over that role in comparison with other users. The user’s priority weighting factor is added to the role’s priority level to produce the weighted priority, shown in Figure 2. The weighted priority requires additional rules for controlling the data objects, which are:

1. Users accessing the data object from the same role or a role of the same priority level cannot block the data object from other users.
2. Users accessing the data object from different roles that are not of the same priority level may block the data object from other users, dependent on the difference between the two user’s weighted priorities.

Figure 2 provides a simple example of the priority context. User 1 has higher priority weightings than user 2 and user 3, except for the normal priority level which is equal with user 3. User 2 has higher priority weightings than user 3. Role 1 has a higher priority level than Role 2. Therefore, if user 1 accesses the data object with Role 1, the user will have higher priority access than user 2 and may be able to block user 3 from the data object. However, if user 1 accesses the data object with Role 2, user 2 may be able to block the data object from both user 1 and user 2. User 1 will have the same priority access privileges to the data object as user 3, which will be handled on a first come first served basis.

3.2 LOCALITY

The utility devices may be accessed remotely or locally by different stakeholders with conflicting operations, and as a result safeguards are required to protect local users from improper remote user operations, e.g. preventing a remote user from turning off an electric water heater in winter. Therefore, a local user may be able to preempt a remote user’s operations even though the remote user’s weighted priority is greater than the local user’s weighted priority. The determination of when a local user’s priority exceeds a remote user’s priority is device dependent. However, the determination of the local versus remote user’s operational priority is made systematic by extending the priority context, as shown in Figure 3.
In Figure 3, a user is assigned a local access role and a remote access role. The local access role is typically one priority level higher than the remote access role. The local access role therefore gives a higher weighted priority than for the remote users, allowing the local access role to block remote users from controlling the data object. To reduce the amount of supported roles, the RBAC infrastructure concept of parameters differentiating roles among devices also includes the priority level of the role. A user’s remote and local roles therefore may be exactly the same except for the priority level of the role, and therefore also the user’s priority weighting factor for that role. A user may either use a local role or a remote role but not both simultaneously, creating mutual exclusion to ensure that user security is not compromised and operation conflicts do not occur.

The priority parameters for differentiating roles can complicate RBAC inheritance since the priority weighting would be the result of multiple role parameter interactions. Therefore, the RBAC infrastructure uses the concept of master roles and sub-roles to handle the parameters systematically. The master roles interact with the individual users, and contain the role’s parameters that dictate the roles’ overall behaviour. The sub-roles are role modules inherited by the master roles for specific operations, such as a DNP3 authentication security role for each user. The sub-roles are not accessible by users, as shown in Figure 4. The sub-roles are handled as universal roles, thereby reducing the amount of roles required to be supported or created by the manufacturer.

![Figure 4: RBAC Infrastructure Role Inheritance](image)

### 3.3 Concurrency

The SCADA devices are operated by multiple stakeholders and therefore device operation conflicts may arise. The priority and locality contexts may not be sufficient to resolve these conflicts since these contexts do not handle the fundamental right to control the data object. Therefore the RBAC infrastructure utilizes role concurrency to determine when control operations are allowed by the users, and is similar to separation of duty [9]-[10]. In the RBAC concurrency context example shown in Figure 5, the two users assigned to two different roles must both set control flags before either role can control the data object, e.g. a consumer providing permission to the utility to turn off the air-conditioning during peak demand.

![Figure 5: RBAC Infrastructure Concurrency Context](image)

### 3.4 Temporal

The priority, locality, and concurrency RBAC contexts may not be sufficient to resolve all user conflicts resulting from multiple stakeholders. Temporal context constraints create finer control granularity for roles [12], [16]-[18] by recognizing situations where the timing of the device operations is as critical as the operation itself. The temporal constraints place limitations on when a role is allowed to be used and the degree of control a role has over the operations in combination with the other contexts. An example of temporal constraints applied to roles is shown in Figure 6.

![Figure 6: RBAC Infrastructure Temporal Context](image)
In Figure 6, Role 1 can only be used during certain hours of the day, e.g. disabling a user’s remote access to the SCADA device when the user is at home. Role 2 ensures that the role can only be used for a specific duration, e.g. a utility restricted in the amount of time that air conditioning can be turned off. Role 3 ensures that the role capabilities are enabled or disabled for specific durations, e.g. preventing users assigned to the same role with the same priority rapidly issuing conflicting commands.

4. DEMAND SIDE MANAGEMENT EXAMPLE

A demand side management example is used to illustrate the RBAC infrastructure presented in this paper. The DSM in this example is used for metering, large appliance control (air conditioning), generation control, and grid disconnects. The DSM operations are controlled through a server within the residence, shown in Figure 7. The server connects the users locally, e.g. via Ethernet, and remotely. Only one user is defined per utility, electricity retailer, and government stakeholders. The consumer stakeholder is divided into the primary consumer role (P Consumer), e.g. parents, and the secondary consumer role (S Consumer), e.g. children.

Role 1 in Figure 8 accesses the electricity monitoring operations. Role 2 is used for disconnecting the power grid to the home. Role 3 is used for the electricity rate operations. Role 4 is used to turn on and off high demand appliances, such as air conditioning. Role 5 is used for the high demand appliance load shaping, e.g. lowering the air conditioning. Role 6 is used for disconnecting the local ‘green’ generation from the grid. Role 7 controls the operations for the ‘green’ generation including the amount of electricity sent into the grid.
Table 1 provides the user priority weighting factors for each priority level. In this DSM example, the low priority level is used for monitoring operations and therefore the weighting factors are used to only to determine the order of preferential treatment for the stakeholders. The normal priority level is used for routine DSM operations. This priority level also corresponds to the primary consumer’s remote access for most roles. The utility has better situational awareness than the consumer’s remote access, and therefore the utility is given a higher weighting factor than the consumer. The government is given a lower weighting factor than the consumer since the government has less need to access the DSM operations. The high priority level is used for remote access to system controls as well as for the primary consumer’s local access. As a result, the primary consumer is given a higher priority weighting factor than the utility since the consumer will have a better local situational awareness than the utility.

Table 1: User Priority Weighting Factor Ranges for the Role Priority Levels

<table>
<thead>
<tr>
<th>Role</th>
<th>Low Priority Level</th>
<th>Normal Priority Level</th>
<th>High Priority Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>Low base value +2</td>
<td>Normal base value +4</td>
<td>High base value +5</td>
</tr>
<tr>
<td>Retailer</td>
<td>Low base value +0</td>
<td>Normal base value +0</td>
<td>High base value +0</td>
</tr>
<tr>
<td>Utility</td>
<td>Low base value +4</td>
<td>Normal base value +6</td>
<td>High base value +8</td>
</tr>
<tr>
<td>Primary Consumer</td>
<td>Low base value +3</td>
<td>Normal base value +5</td>
<td>High base value +9</td>
</tr>
<tr>
<td>Secondary Consumer</td>
<td>Low base value +1</td>
<td>Normal base value +1</td>
<td>High base value +3</td>
</tr>
</tbody>
</table>

Table 2 shows the remote access priority level for each role. Secondary consumers (S Consumer) do not require remote access to or control over the DSM operations and therefore only have local roles. For Role 1, reading the DSM measurement values are low priority, e.g. power consumption readings are typically measured every 15 minutes [12]. The utility disconnect, Role 2, may be utilized for safety considerations, such as fuse and breaker failures, and is therefore high priority. The government’s weighted priority is lower than the utility and consumer for Role 2 since government intervention would be more likely due to utility and consumer disputes rather than safety considerations. Role 3, electricity rate settings, is a revenue operation that must be timely and is therefore assigned to the normal priority level. The utility’s weighted priority is higher for Role 3 than the retailer for peak demand management, such as using aggressive high pricing to discourage power consumption. Role 4 is a high priority role in order to handle power system problems that the utility is more responsive to than the consumer. Role 5 is used for typical DSM management of the high electricity demand appliances. The utility has a slight priority edge over the consumer for this role since the consumer is accessing DSM remotely. Role 6 is used for emergency generation disconnects to the grid, e.g. due to servicing or fault restoration. This role is oriented more towards the utility than the consumer, and therefore the utility is given a higher priority. For Role 7, the utility has a slight priority edge over the consumer for this role since the consumer is accessing DSM remotely.

Table 2: Remote Access Role Priority Level

<table>
<thead>
<tr>
<th>Role</th>
<th>Role 1</th>
<th>Role 2</th>
<th>Role 3</th>
<th>Role 4</th>
<th>Role 5</th>
<th>Role 6</th>
<th>Role 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>Low</td>
<td>High</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Retailer</td>
<td>Low</td>
<td>-</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Utility</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Normal</td>
<td>High</td>
<td>Normal</td>
</tr>
<tr>
<td>Primary Consumer</td>
<td>Low</td>
<td>High</td>
<td>-</td>
<td>Normal</td>
<td>Normal</td>
<td>Low</td>
<td>Normal</td>
</tr>
<tr>
<td>S Consumer</td>
<td>Low</td>
<td>High</td>
<td>-</td>
<td>Normal</td>
<td>Normal</td>
<td>Low</td>
<td>Normal</td>
</tr>
</tbody>
</table>

In Table 3, the local access priority levels are given for the utility and consumer. For the local access, the primary consumer is given the priority edge over the utility’s local and remote access since the primary consumer has increased local situational awareness and increased control requirements, e.g. greater determination of residential comfort levels. The utility retains higher priority control over the grid disconnection operations, Role 2 and Role 6, for handling emergency conditions. The secondary consumer is given control over the local environment only and not the generation to prevent them from altering the primary consumer’s revenue sources. The secondary consumer has an overall priority less than the utility and the primary consumer, which allows the primary consumer to block the secondary consumer from controlling other roles that are not described here, e.g. blocking the secondary consumer from accessing various device attributes. In addition, the difference between the weighted priority for the primary and secondary consumer should be great enough to allow the primary consumer to manage the secondary consumer roles based on the priority context only, e.g. adding or eliminating secondary consumer roles.
Table 3: Local Access Role Priority Level

<table>
<thead>
<tr>
<th>Role 1</th>
<th>Role 2</th>
<th>Role 3</th>
<th>Role 4</th>
<th>Role 5</th>
<th>Role 6</th>
<th>Role 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Retailer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Utility</td>
<td>Low</td>
<td>High</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P Consumer</td>
<td>Normal</td>
<td>High</td>
<td>-</td>
<td>High</td>
<td>High</td>
<td>-</td>
</tr>
<tr>
<td>S Consumer</td>
<td>Normal</td>
<td>-</td>
<td>High</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4 provides the concurrency context for the roles. For Role 2, two of the three stakeholders for this role must be in accordance before the electricity may be disconnected or re-connected at the residence. If the utility and primary consumer are in accordance, the disconnection will be due to either an emergency or the consumer changing residences. If the government is in accordance with one of the other parties, it may be due to an emergency or a revenue dispute between the consumer and the utility, e.g. unpaid bills. For Role 3, the electricity retailer can only change the billing rates with the permission of the utility. For example, this role’s concurrency context can be used to prevent unscrupulous practices by the retailer or to override the retailer rates during special cases, e.g. when demand exceeds capacity and the utility alters the electricity rate to encourage consumers to reduce power consumption rather than using load shedding. For Role 4, the utility requires the permission of both the primary and secondary consumers in order to turn off the high electricity demand devices, e.g. air conditioning. This prevents the utility from creating high discomfort levels for people within the residence. The secondary consumer requires the permission of the primary consumer to turn off the devices, but not to turn them on. For Role 5, the utility requires the permission of both the primary and secondary consumers to change the comfort levels within the home. For Role 6, the consumer requires permission from the utility to re-connect the generation to the grid due to safety concerns, but not to turn the generation off. For Role 7, the utility requires permission from the primary consumer to access the generation control, such as when revenue agreements are reached.

Table 4: Concurrency Role Context

<table>
<thead>
<tr>
<th>Role 1</th>
<th>Role 2</th>
<th>Role 3</th>
<th>Role 4</th>
<th>Role 5</th>
<th>Role 6</th>
<th>Role 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>-</td>
<td>Utility, P Consumer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Retailer</td>
<td>-</td>
<td>-</td>
<td>Utility</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Utility</td>
<td>-</td>
<td>Government, P Consumer</td>
<td>-</td>
<td>Consumer</td>
<td>Consumer</td>
<td>-</td>
</tr>
<tr>
<td>P Consumer</td>
<td>-</td>
<td>-</td>
<td>Utility</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S Consumer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>P Consumer</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In Table 5, the temporal contexts are given. The electricity retailer is prevented from altering the electricity rate more than once an hour using Role 3, thereby avoiding fraudulent pricing practices. The utility is prevented from altering the DSM settings during the hours when the family is normally at home, or if more relaxed rules are used, during typical sleep periods. The primary consumer is prevented from using the remote access roles during typical local access periods. This is more than simply exclusion, but the prevention of remote access if the local access is not being utilized. For the remote access roles, the primary consumer is prevented from making device setting changes for a specific duration after another primary user changes the settings. The primary consumer remote access roles are prevented from drastically altering the homes’ comfort level (Role 4 and Role 5) when the secondary consumers are typically present within the residence, e.g. upper and lower control limits are dictated.

Table 5: Temporal Role Context

<table>
<thead>
<tr>
<th>Role 1</th>
<th>Role 2</th>
<th>Role 3</th>
<th>Role 4</th>
<th>Role 5</th>
<th>Role 6</th>
<th>Role 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Retailer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Utility</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✔</td>
</tr>
<tr>
<td>P Consumer</td>
<td>✔</td>
<td>✔</td>
<td>-</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>S Consumer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
4. SUMMARY

Aggressive DSM is required to fully engage consumers in actively reducing or shaping their load demands. This aggressive DSM must be supported with remote access by multiple stakeholders, including the utility, consumer, government, and the electricity retailers. The multiple stakeholders create data access and control conflicts that can be resolved through RBAC. The RBAC presented in this paper utilizes power system characteristics to limit the effect of RBAC on SCADA devices, which typically have limited processing capability, by reducing the amount of users and roles required. The reduction of the users and roles is achieved through the two-tier model, the use of parameter based roles, and the master roles and sub-roles. Reducing the amount of users and roles for the SCADA devices increases data object access and control conflicts. These conflicts are resolved with the priority, locality, concurrency, and temporal contexts. The priority context determines which user has priority access to the data. The priority context also determines the capability to block the data object from other user access or control. The locality context differentiates between local and remote users, and is used to determine which user has the greater situational awareness of the DSM operations. Concurrency ensures accordance among users, preventing one user from exercising unauthorized control over DSM operations. The temporal context blocks role access to data objects at specific times or for a specific duration, which prevents users from abusing the SCADA device or from increasing discomfort levels within the home.

REFERENCES

Virtual Enterprise Procurement based on Combinatorial Reverse Auction

Fu-Shiung Hsieh

Department of Computer Science and Information Engineering
Chaoyang University of Technology
Taichung County, 41349 Taiwan

ABSTRACT

Creation of a virtual enterprise (VE) is often driven by the lack of capabilities to fulfill the customers’ order requirements by a single firm. In case the requirements of an order cannot be met by a single firm, a VE may be formed via procurement of goods or services from partners to respond to the business opportunity. Procurement of goods or services is often accomplished by reverse auctions, where suppliers provide their competitive biddings to a buyer. Combinatorial reverse auction can be applied in procurement to purchase goods at the lowest possible cost. A buyer can hold a reverse auction to try to obtain the goods from a set of sellers who places bids for each bundle of goods he can provide. The objectives of this paper are to: (1) propose architecture for procurement in VE based on combinatorial reverse auction, (2) formulate combinatorial reverse auction problem to minimize the cost of a VE and (3) develop a solution algorithm to find a near-optimal solution efficiently. We consider a winner determination problem for combinatorial reverse auction in which a buyer wants to acquire items from a set of sellers and each seller can provide a set of items. The problem is to determine the winners to minimize the total cost to acquire the required items. The main results include: (1) a problem formulation for the combinatorial reverse auction problem; (2) a solution methodology based on Lagrangian relaxation; (3) a heuristic algorithm for finding a near-optimal feasible solution and (4) results and analysis of our solution algorithms.

1. INTRODUCTION

In a virtual enterprise (VE) [1], a company assembles a temporary consortium of partners and services for a customer’s request, fulfillment orders, or taking advantage of a new resource or market niche. The general rationale for forming a VE is to reduce costs and time to market while increasing flexibility and access to new markets and resources. Individual companies in a VE focus on their core competencies and mission critical operations and outsource everything else. A wide variety of research issues and topics have been studied, including cooperation/coordination [2], formation [3], partner selection [4],[6], and planning and control [5] of virtual enterprise. Several projects that focus on the study of VE have been launched recently. Among them are the NIIP project in US, the PRODNET project and the VEGA project in Europe. Creation of a successful VE relies on an effective information infrastructure to publish, discover partners and invoke services provided by partners. With the widely deployment of Web services and service-oriented architecture (SOA) [13], service discovery as well as composition can be effectively achieved.

Selection of partners [6] is a key challenge in creation of a VE. Creation of a VE is often driven by the lack of capabilities to fulfill the customers’ order requirements by a single firm. In case the requirements of an order cannot be met by a single firm, a VE may be formed via procurement of goods or services from partners to respond to the business opportunity. Procurement of goods or services is often accomplished by reverse auctions, where suppliers provide their competitive biddings to a buyer [21] and [22]. Each supplier indicates the minimum price at which it is willing to undertake the work or provide the goods/services. Since the assignment is typically awarded to the supplier providing the lowest bid, each supplier is spurred to provide the lowest possible bid, taking into account the expected level of competition and its expected rate of return. Combinatorial reverse auctions are popular, distributed and efficient preserving ways of allocating items or tasks among multiple agents to maximize revenue or minimize cost. In a combinatorial auction, bidders may place bids on combinations of items or tasks. An interesting question is how to
combine SOA and combinatorial reverse auction to enable individual firms to provide their services and consume others’ services on the Web to minimize the cost of a VE. The objectives of this paper are to: (1) propose architecture for procurement in VE based on service-oriented architecture (SOA) [13] and combinatorial reverse auction, (2) formulate combinatorial reverse auction problem to minimize the cost of a VE and (3) develop a solution algorithm to find a near-optimal solution efficiently.

Combinatorial auctions have attracted considerable attention in the existing literature [7]. Combinatorial auctions have been notoriously difficult to solve from a computational point of view [17] due to the exponential growth of the number of combinations [18]. The combinatorial auction problem can be modeled as a set packing problem (SPP) [8],[9],[10],[11],[19]. Sandholm et al. mentions that determining the winners so as to maximize revenue in combinatorial auction is NP-complete [11], [14]. Many algorithms have been developed for combinatorial auction problems. For example, in [12], the authors proposed a Lagrangian Heuristic for a combinatorial auction problem. Exact algorithms have been developed for the SPP problem, including a branch and bound search, iterative deepening A* search [11] and the direct application of available CPLEX IP solver [8]. One way to reduce the computational complexity in solving the Winner Determination Problem (WDP) for combinatorial reverse auction is to apply Lagrangian relaxation technique to develop a solution algorithm for WDP. Lagrangian relaxation provides a systematic approach to determine an allocation and prices based on the introduction of Lagrange multipliers, which set prices for each item to be purchased by the buyer. If two or more sellers compete for the same item, the price will be adjusted. This saves bidders from specifying their bids for every possible combination and the buyer from having to process each bid function. Based on the price for the individual items, bidders submit bids. The bundle associated with a bid is tentatively assigned to that bidder only if the price of the bid is the lowest. Based on the iterative price adjustment mechanism, a solution will be obtained. It should be emphasized that Lagrangian relaxation is not guaranteed to find the optimal solution to the underlying problem. Furthermore, it is not guaranteed to produce a feasible solution by applying Lagrangian relaxation technique. In case the resulting solution is not feasible, a heuristic algorithm must be applied to adjust the infeasible solution to a feasible one. We develop a heuristic algorithm for finding a near-optimal, feasible solution based on the solution of the relaxed problem.

This paper is differentiated from [2],[3],[4],[5] in that we focus on the optimization aspect in formation of VE. This paper is different from the one presented in [12] as we adopt the subgradient method find Lagrange multipliers. The remainder of this paper is organized as follows. In Section 2, we present architecture of combinatorial reverse auction for VE. In Section 3, we formulate the combinatorial reverse auction problem. In Section 4, we propose a solution approach for the combinatorial reverse auction based on Lagrange relaxation. In Section 5, we present experimental results and analysis. We conclude this paper in Section 6.

2. COMBINATORIAL REVERSE AUCTION FOR VIRTUAL ENTERPRISES

Given an order with specific product demands and due date, the problem is to form a virtual enterprise for the order and determines whether the order can be committed based on the available resources in a VE. Consider a scenario in which a customer places an order to a reseller. Suppose the order requirements include a bundle of items and the reseller is not able to provide all the items on its own. In this case, the reseller will try to acquire the required items from the suppliers or partners. Service-oriented architecture (SOA) provides an infrastructure to discover partners and services to process an order. However, SOA does not provide the mechanism to determine the best partners. Auctions are popular, distributed and autonomy preserving ways of allocating items or tasks among multiple agents to maximize revenue or minimize cost. In economics, different types of auctions have been proposed and extensively studied, including English Auction (open ascending price auction), Dutch auction (open descending price auction) and sealed first-price auction, etc. Single item auctions are by far the most common auction format, but they are not always efficient. Combinatorial auctions [14], [15], [16] enable several bidders to bid on different combination of goods according to personal preferences in the auction process. Allowing bids for bundles of items is the foundation of combinatorial auctions. Bidders can select multiple items at one time and offer those items a price. It enables bidders to decide combinations of auction according to personal preferences of bidders. Combinatorial auctions are beneficial if complementarities exist between the items to be auctioned.

To endow the SOA infrastructure with the capability to determine the best partners, we propose architecture based on combinatorial reverse auction as shown in Figure 1, where five steps are involved to form VE.

Step 1: Each seller registers their products/services in the registry.
Step 2: Each buyer looks up the registry to find the potential sellers.
Step 3: After finding the potential sellers, the buyer announces a CFP (Call For Proposal) message to request the potential sellers for tender.

Step 4: The potential sellers draw up proposals and place bids by sending a SOP (Submission of Proposal) message to the buyer.

Step 5: The buyer determines the winners and awards the contracts to the winners.

Figure 1: Virtual Enterprises Procurement in SOA

Figure 2: Combinatorial Reverse Auction

The reseller may hold a combinatorial reverse auction to minimize the cost. Figure 2 illustrates an application scenario in which Reseller requests to purchase at least a bundle of items 2A, 3B, 2C and 1D from the market. There are three bidders, Supplier 1, Supplier 2 and Supplier 3 who place bids in the system. Suppose Supplier 1 places two bids: (1A, 2B, p11) and (1C, 1D, p12), where p11 and p12 denote the prices of the bids. Supplier 2 places two bids: (1B, 2C, p21) and (2C, 1D, p22). Supplier 3 places two bids: (1C, 1D, p31) and (1A, 1B, p32). We assume that all the bids entered the auction are recorded. A bid is said to be active if it is in the solution. We assume that there is only one bid active for all the bids placed by the same bidder. For this example, the solution for this reverse auction problem is Supplier 1: (1A, 2B, p11), Supplier 2: (2C, 1D, p22) and Supplier 3: (1A, 1B, p32).

3. PROBLEM FORMULATION

Combinatorial reverse auction can be applied in procurement to purchase goods at the lowest possible cost. A buyer can hold a reverse auction to try to obtain the goods from a set of sellers who can provide the goods. Each seller places bids for each bundle of goods he can provide. Although combinatorial reverse auctions have attracted a lot of attention recently, most studies focus on development of efficient solution algorithms to determine winners. From the viewpoint of a buyer, an important issue is to design an effective algorithm to collectively minimize the overall cost. In this paper, we first formulate the above combinatorial optimization problem as an integer programming problem. We then develop solution algorithms based on Lagrangian relaxation.

Consider a buyer who requests a set of items to be purchased. Let \( K \) denote the number of items requested. Let \( d_k \) denote the desired units of the \( k \)-th items, where \( k \in \{1, 2, 3, \ldots, K\} \). In a combinatorial auction, there are many bidders to submit a tender. Let \( I \) denote the number of bidders in a combinatorial auction. Each \( i \in \{1, 2, 3, \ldots, I\} \) represents a bidder. To model the combinatorial auction problem, the bid must be represented mathematically. We use a vector \( b_{ij} = (q_{ij1}, q_{ij2}, q_{ij3}, \ldots, q_{ijk}, p_{ij}) \) to represent the \( j \)-th bid submitted by bidder \( i \), where \( q_{ijk} \) is a nonnegative integer that denotes the quantity of the \( k \)-th items and \( p_{ij} \) is a real positive number that denotes the price of the bundle. As the quantity of the \( k \)-th items cannot exceed the quantity \( d_k \), it follows that the constraint \( 0 \leq q_{ijk} \leq d_k \) must be satisfied. The \( j \)-th bid \( b_{ij} \) is actually an offer to deliver \( q_{ijk} \) units of items for each \( k \in \{1, 2, 3, \ldots, K\} \) a total price of \( p_{ij} \). Let \( n_i \) denote the number of bids placed by bidder \( i \in \{1, 2, 3, \ldots, I\} \). To
formulate the problem, we use the variable \( x_{ij} \) to indicate the \( j \)th bid placed by bidder \( i \) is active (\( x_{ij} = 1 \)) or inactive (\( x_{ij} = 0 \)). The winner determination problem can be formulated as an Integer Programming problem as follows.

**Winner Determination Problem (WDP):**

\[
\begin{align*}
\min & \sum_{i=1}^{I} \sum_{j=1}^{n_i} x_{ij} p_{ij} \\
\text{s.t.} & \sum_{i=1}^{I} \sum_{j=1}^{n_i} x_{ij} q_{ijk} \geq d_k \quad \forall k = 1, 2, \ldots, K \\
& \sum_{i=1}^{I} x_{ij} \leq 1 \quad \forall i = 1, \ldots, I \\
& x_{ij} \in \{0,1\}
\end{align*}
\]

In WDP, we observe that the coupling among different operations is caused by the contention for resources through the minimal resource requirement constraints (2-1).

To find a solution to WDP requires the development of combinatorial reverse auction algorithms. By applying an effective combinatorial reverse auction algorithm, the reseller will be able to optimize the overall costs.

Development of an effective combinatorial reverse auction algorithm to solve WDP is a key issue. Combinatorial auctions have attracted considerable attention in the existing literature. An excellent survey on combinatorial auctions can be found in [7] and [16]. Combinatorial auctions have been notoriously difficult to solve from a computational point of view [17] due to the exponential growth of the number of combinations [18]. The combinatorial auction problem can be modeled as a set packing problem (SPP) [8],[9],[10], [19]. Sandholm et al. mentions that determining the winners so as to maximize revenue in combinatorial auction is NP-complete [11], [14]. Many algorithms have been developed for combinatorial auction problems.

One way to reduce the computational complexity in solving the Winner Determination Problem (WDP) for combinatorial reverse auction is to set up a fictitious market to determine an allocation and prices in a decentralized way to adapt to dynamic environments where bidders and items may change from time to time. In this paper, we apply Lagrangian relaxation technique to develop a solution algorithm for WDP.

Lagrangian relaxation provides a systematic approach to determine an allocation and prices based on the introduction of Lagrange multipliers, which set prices for each item to be purchased by the buyer. If two or more sellers compete for the same item, the price will be adjusted. This saves bidders from specifying their bids for every possible combination and the buyer from having to process each bid function. Based on the price for the individual items, bidders submit bids. The bundle associated with a bid is tentatively assigned to that bidder only if the price of the bid is the lowest. Based on the iterative price adjustment mechanism, a solution will be obtained. It should be emphasized that Lagrangian relaxation is not guaranteed to find the optimal solution to the underlying problem. Furthermore, it is not guaranteed to produce a feasible solution by applying Lagrangian relaxation technique. In case the resulting solution is not feasible, a heuristic algorithm must be applied to adjust the infeasible solution to a feasible one. We develop a heuristic algorithm for finding a near-optimal, feasible solution based on the solution of the relaxed problem.

**4. Solution Algorithm**

For a given Lagrange multiplier \( \lambda \), the relaxation of constraints (2-1) decomposes the original problem into a number of bidder’s subproblems (BS). These subproblems can be solved independently. That is, the Lagrangian relaxation results in subproblems with a highly decentralized decision making structure. Interactions among subproblems are reflected through Lagrange multipliers, which are determined by solving the following dual problem.

\[
\max_{\lambda \geq 0} L(\lambda), \text{ where } \\
L(\lambda) = \sum_{k=1}^{K} \lambda_k d_k + L_1(\lambda), \text{ with } L_1(\lambda) = \min \sum_{j=1}^{n_i} x_{ij} (P_{ij} - \sum_{k=1}^{K} \lambda_k q_{ijk}) \\
\text{s.t.} \sum_{j=1}^{n_i} x_{ij} \leq 1, x_{ij} \in \{0,1\}
\]
Virtual Enterprise Procurement based on Combinatorial Reverse Auction

A subgradient method for solving the dual problem

Let \(\alpha^l\) be the optimal solution to the subproblems for given lagrange multipliers \(\lambda^l\) of iteration \(l\). We define the subgradient of \(L(\lambda)\) as

\[
g_k^l = \frac{\partial L(\lambda)}{\partial \lambda_k} = d_k - \sum_{i=1}^{l} \sum_{j=1}^{n_i} x_{ij} q_{ijk}, \text{ where } k \in \{1,2,...,K\}.
\]

The subgradient method proposed by Polak [9] is adopted to update \(\lambda\) as follows

\[
\lambda_k^{l+1} = \begin{cases} 
\lambda_k^l + \alpha^l g_k^l & \text{if } \lambda_k^l + \alpha^l g_k^l \geq 0; \\
0 & \text{otherwise.}
\end{cases}
\]

where \(\alpha^l = c \frac{L - L(\lambda)}{\sum (g_k^l)^2} \), \(0 \leq c \leq 2\) and \(L\) is an estimate of the optimal dual cost. The iteration step terminates if \(\alpha^l\) is smaller than a threshold. Polyak proved that this method has a linear convergence rate.

Iterative application of the algorithms in (1) and (2) may converge to an optimal dual solution \((x^*, \lambda^*)\).

(3) A heuristic algorithm for finding a near-optimal, feasible solution based on the solution of the relaxed problem

The solution \((x^*, \lambda^*)\) may result in one type of constraint violation due to relaxation: assignment of the quantity of items less than the demand of the items. Our heuristic scheme first checks all the demand constraints

\[
\sum_{i=1}^{l} \sum_{j=1}^{n_i} x_{ij} q_{ijk} \geq d_k \quad \forall k = 1,2,...,K.
\]

Let \(K^0 = \{k | k \in \{1,2,3,...,K\}, \sum_{i=1}^{l} \sum_{j=1}^{n_i} x_{ij} q_{ijk} < d_k\}\). \(K^0\) denotes the set of demand constraints violated.

Let \(l^0 = \{l | l \in \{1,2,3,...,l\}, x_{ij}^* = 0\}\). \(l^0\) denotes the set of bidders that is not a winner in solution \(x^*\). To make the set of constraints \(K^0\) satisfied, we first pick \(k \in K^0\) with \(k = \arg \min_{k \in K^0} d_k - \sum_{i=1}^{l} \sum_{j=1}^{n_i} x_{ij}^* q_{ijk}\). The heuristic algorithm proceeds...
as follows to make constraint \( k \) satisfied. Select \( i \in I^0 \) and \( j \in \{1, 2, \ldots, n_i\} \) with \( j = \arg \min_{j \in \{1, 2, \ldots, n_i\}, q_{ij} > 0} p_{ij} \) and set \( x_{ij}^* = 1 \).

After performing the above operation, we set \( I^0 \leftarrow I^0 \setminus \{i\} \). If the violation of the \( k \)-th constraint cannot be completely resolved, the same procedure repeats. Eventually, all the constraints will be satisfied.

5. Experimental Results and Analysis

Based on the proposed algorithms for combinatorial reverse auction, we conduct several examples to illustrate the effectiveness of our method.

Example 1: Consider a buyer who will purchase a set of four items. The desired units of each type of items are
\[
\begin{align*}
d_1 &= 2, \\
d_2 &= 1, \\
d_3 &= 2, \\
d_4 &= 1.
\end{align*}
\]
Suppose there are three sellers. Suppose each bidder only places two bids. For this example, we have
\[
I = 3, \quad J = 2, \quad K = 4, \\
d_1 &= 2, d_2 = 1, d_3 = 2, d_4 = 1.
\]
Suppose the four bids submitted by the two bidders are as follows:
\[
\begin{align*}
q_{111} &= 1, q_{112} = 0, q_{113} = I, q_{114} = 0, \\
q_{121} &= I, q_{122} = 1, q_{123} = 0, q_{124} = 0, \\
q_{211} &= 0, q_{212} = 0, q_{213} = I, q_{214} = 0, \\
q_{221} &= 0, q_{222} = I, q_{223} = 0, q_{224} = I, \\
q_{311} &= 0, q_{312} = 0, q_{313} = I, q_{314} = 0, \\
q_{321} &= 0, q_{322} = 0, q_{323} = 0, q_{324} = 1.
\end{align*}
\]
Suppose we initialize the Lagrange multipliers as follows.
\[
\lambda(1) = 50.0, \lambda(2) = 40.0, \lambda(3) = 35.0, \lambda(4) = 50.0.
\]
Our algorithm the subgradient algorithm converges to the following solution:
\[
\begin{align*}
x_{12}^* &= 1, \\
x_{22}^* &= 1, \\
x_{32}^* &= 1.
\end{align*}
\]
As the above solution is not a feasible one, the heuristic algorithm needs to be applied. Our heuristic algorithm leads to the following feasible solution
\[
\begin{align*}
X_{11} &= 1, \\
X_{21} &= 1, \\
X_{22} &= 1, \\
X_{32} &= 1.
\end{align*}
\]
The duality gap of the solution is as follows:
Duality gap = 2.448%.
Despite the duality gap is not zero, the solution \( X \) is so an optimal solution for this example.
In addition to Example 1, Table 1 illustrates the duality gap of several cases based on the problem size (I, J, K). According the results, the duality gaps are within 3%. This means the solution methodology generates near-optimal solution.

<table>
<thead>
<tr>
<th>I</th>
<th>J</th>
<th>K</th>
<th>Duality Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>5</td>
<td>2.8%</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>2.2%</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>20</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Table 1: Duality Gap of Several Cases
In addition to the two examples above, we also conduct several experiments to study the computational efficiency of our proposed algorithms. These experiments shows the growth of CPU time with respect to $I$, $J$, and $K$, respectively. Figure 3 shows the CPU time for a number of problems in which parameter $J$ and $K$ are fixed while the parameter $I$ is changed. The CPU time for a number of problems in which parameter $I$ and $K$ are fixed while the parameter $J$ is changed is shown in Fig. 4. Figure 5 shows the growth of CPU time with respect to $K$.

6. CONCLUSION

Creation of a virtual enterprise (VE) is usually driven by the customers’ orders. When an order arrives, one must determine whether a VE can be formed to fulfill the order requirements. In case the order requirements cannot be met by a single firm, a VE may be formed via procurement of goods or services from partners to respond to the business opportunity. Procurement of goods or services is often accomplished by reverse auctions, where suppliers provide their competitive biddings to a buyer. Combinatorial reverse auctions are popular, distributed and autonomy preserving ways of allocating items or tasks among multiple agents to minimize the cost of a VE. We propose architecture for forming a VE based on service-oriented architecture and combinatorial reverse auction.

Combinatorial reverse auctions enable several bidders to bid on different combination of goods simultaneously according to personal preferences. Bidders can select multiple items at one time and offer those items a combined price. It enables bidders to decide combinations of auction according to personal preferences and more effectively arranges bid winners. Combinatorial reverse auctions have been notoriously difficult to solve from a computational point of view due to the exponential growth of the number of combinations. We formulate the combinatorial reverse auction problem and propose a combinatorial reverse auction algorithm based on a Lagrange relaxation approach. We formulate a winner determination optimization problem for combinatorial reverse auction. The demands of the buyer impose additional constraints on determination of the winners. By applying Lagrangian relaxation technique, the original optimization can be decomposed into a number of bidders’ subproblems. Our methodology consists of three parts: (1) an algorithm for solving bidders’ subproblems by exploiting their individual structures; (2) a subgradient method for solving the non-differentiable dual problem; (3) an economic interpretation of Lagrange multipliers; (4) a heuristic algorithm for
finding a near-optimal, feasible solution and (5) results and analysis of our solution algorithms. Numerical results indicate that our proposed algorithms yield near optimal solutions.

ACKNOWLEDGEMENTS

This paper is supported in part by National Science Council under Grant NSC97-2410-H-324-017-MY3.

REFERENCES


Integration of Product Design, Process Planning and FMS Control using Neutral Data Representation

Dusan N. Sormaz¹, Chintan Patel

Department of Industrial and Systems Engineering, Ohio University, Athens, OH 45701, USA

ABSTRACT

An efficient model for communications between CAD, CAPP and CAM applications in distributed manufacturing planning environment has been seen as a key ingredient for CIM. Integration of design model with process and scheduling information in real-time is necessary in order to increase product quality, reduce the cost and shorten the product manufacturing cycle. The paper describes an approach to integrate key product realization activities using neutral data representation. The representation is based on established standards for product data exchange and serves as a prototype implementation of these standards. The product and process models are based on object-oriented representation of geometry, features, and resulting manufacturing processes. Relationships between objects are explicitly represented in the model (for example, feature precedence relations, process sequences, etc.). The product model is developed using XML-based representation for product data required for process planning and the process model also uses XML representation of data required for scheduling. The procedures for writing and parsing XML representations have been developed in object-oriented approach, in such a way that each object from object oriented model is responsible for storing its own data into XML format. Similar approach is adopted for reading and parsing of the XML model. Parsing is performed by a stack of XML handlers, each corresponding to a particular object in XML hierarchical model. This approach allows for very flexible representation, in such a way that only a portion of the model (for example, only feature data, or only the part of process plan for a single machine) may be stored and successfully parsed into another application. This is very useful approach for direct distributed applications, in which data are passed in the form of XML streams to allow real-time on-line communication. The feasibility of the proposed model is verified in a sample scenario for distributed manufacturing planning that involves feature mapping from CAD file, process selection for several part designs and simulation of the FMS model using alternative routings.

1. INTRODUCTION

Computer Integrated Manufacturing (CIM) is still moving target in industrial and manufacturing engineering research and applications. There are numerous reports outlining individual islands of automation within the CIM model, with a significant effort in research papers being focused on a particular task, and much less effort devoted to integration of these manufacturing engineering and planning tasks. Integration involves the transfer of data between applications, but also should focus on data and model integrity, distributed processing of the data, incorporation of knowledge into the planning tasks, and so on. Computer aided process planning (CAPP) is rightly seen as an integration fabric for CIM with its relations to design (CAD), manufacturing (CAM), and scheduling tasks. By virtue of being an integrator between CAD, CAM, and scheduling, process planning involves decision making process on various levels of details with the final goal of generating feasible and/or optimal process plan. The need for integration requires that such model is transparent between these tasks and that it is easily saved or transferred.

This paper proposes such neutral data model in the form of XML model and describes its details and application in manufacturing process planning. The paper is organized as follows. Section 2 describes previous work in process plan representation and modeling. Section 3 describes process planning representation object model, which includes major entities and relations between them. Section 4 explains process planning XML representation that is built for previous object model and describes methods for writing data in XML format and parsing XML streams into object model. Section 5 describes a sample scenario of integration between feature mapping, process selection and FMS control.
control using data in XML format and section 6 describes a case study that was performed to verify the approach. The paper ends with conclusions and the list of references.

2. **PREVIOUS WORK**

Data and knowledge representation in process planning have received significant interest in research. An early work on ALPS [1] proposed a graphical representation for manufacturing processes and means for specifying serial, parallel and concurrent tasks. Since then, several papers addressed knowledge representation, for example, using frames and rules [2] or an object-oriented data model [3]. International standard for product data exchange (STEP) has been developed to enable data transfer between applications that includes process data [4]. Recent results are in generation of the Process Specification Language (PSL) [5] as a neutral format for the specification of process representation and exchange of different ontologies or semantics between various domains. Development of process planning specific, NC data within STEP standard has been described in [6]. Work on XML [7], as a very flexible language that transfers both data and their description (metadata) prompted its widespread use in many research efforts. Paper [8] discusses similarities and differences between STEP and XML and proposes their convergence.

3. **PROCESS PLANNING REPRESENTATION AND MODELING**

Manufacturing planning object model developed in this work is based on process planning object model proposed in [3], which described a data model for representation of process plans based on the different activities involved in manufacturing. A process planning representation model facilitates the development of algorithms for manufacturing problems like sequencing, scheduling etc. by reducing the over-all algorithm development time. The model accommodates a variety of data that may be needed in manufacturing planning algorithms. Components of the model are manufacturing planning model, feature object model, and process object model and they are described in this section.

3.1 **FEATURE OBJECT MODEL**

The feature object model represents a hierarchical representation of various machining features with inheritance relations within it. The model is shown in Figure 1. The major class is \textit{MfgFeature} that abstracts all common properties for all features. Properties at this level include feature name, containing part, tolerance data, list of alternative processes, and precedence relations. \textit{MfgFeature} class is extended into several subclasses that correspond to machining feature types found in mechanical prismatic parts (such as \textit{Hole}, \textit{Slot}, and \textit{Pocket}). These classes model properties of particular feature type and include different dimension parameters, and process capability data. However, model properties on general, feature level are of generic nature and can be applied by extending this model to other domains (like rotational parts, sheet metal parts, and so on).

![Figure 1. Class diagram for features](image-url)


### 3.2 Machining Process Object Model

The knowledge about processes is also represented in an object-oriented model. In order to map this knowledge representation, the machining processes are categorized based on their characteristics. Figure 2 presents this hierarchy in a UML based model.

The class *MfgProcess* represents the most generic process class, i.e., a process with the common data shared by the rest of the processes. Further distinctions are carried out based on the process characteristics such as hole making processes and profile generating milling processes. The hole making processes are further divided into *CoreMaking*, *HoleStarting* and *HoleImproving* processes, while the *Milling* process has as sub-type *EndMilling* process. These generalized classes are implemented as abstract classes and are shown with italicized titles in Figure 2. The classes under these umbrella classes are for representing the actual machining processes (for example, *TwistDrilling*, *EndMillingSlotting*, or *FaceMilling*). Therefore, based on the inheritance, *EndMillingSlotting* process acquires process information from *EndMilling*, which further leads to the parent class *MfgProcess*. The following paragraphs give a brief description about the *MfgProcess* and *EndMilling* classes., while description of other classes is part of working documentation.

**Figure 2. Machining process hierarchy**

*MfgProcess*: This class contains member variables such as feature, stock, workpiece, cutting parameters, constraints, tool, and tool path. These listed variables are used in every inherited process class as every process contains these components of machining process. Also, this class carries the GUI components for showing process information and a graphical interface to display a process.

*EndMilling*: The implementation for visualization is mainly concentrated on end milling operations, so this class contains some graphical components. This class provides methods to prepare the scene graph, which carries nodes to display machining process components and the animation for visualization. The GUI required for event-based interaction with the virtual world displayed is provided in this class. The subclasses --slotting and peripheral end milling-- use this implementation with distinctive modifications in tool approach for machining.
4. PROCESS PLANNING XML REPRESENTATION

This section discusses the persistent and neutral data storage mechanism for the process plan representation model (PPRM) based on XML. It describes the XML schema which is the basis or template for process plan data generation, provides few examples of XML data, and describes methods for writing and parsing data in XML format.

4.1 XML SCHEMA MODEL

An XML schema is a set of rules that have to be satisfied in valid XML document [7]. Figure 3 shows the XML Schema definition for the process plan representation model as shown in a tool for XML development _XMLSpySuite Error! Reference source not found._.

![XML schema for FMS in XML SpySuite](image)

The root element of the schema is _MfgSystem_ that holds all information about a FMS model. This element holds two collection elements _partsList_ and _machinesList_. The _partsList_ element is a collection of part models represented by the _MfgPartModel_ element. The _machinesList_ is a collection of the available machines represented by _Machine_ element. Each _MfgPartModel_ element has a list element called _partActivityList_, which is a collection of alternative process plans, for the enclosing part element. These alternative process plans are represented as _PartActivity_ elements on the part model. Each _PartActivity_ element has a _machineActivityList_ element. The _machineActivityList_ element is a collection of all the _MachineActivity_ elements in the enclosing process plan. Each _MachineActivity_ element encloses a _Machine_ element that holds information about the machine on which the machining operation is done. Each _MachineActivity_ element holds a _toolDirectionActivityList_ element. This element holds the spatial information about the tools used in this machining activity. _toolDirectionActivity_ in turn, holds _mfgProcessName_ element as reference to a machining process. The hierarchical relationship between parts, machines, part activities, machine activities, tool activities, features and processes is in such way incorporated into the XML document.

Information on features and processes may be included in few levels of detail. For feature modeling, geometric reasoning and process selection all details are needed, so elements actually carry the name of the feature or process class, with inclusion of all details. Example of this model for two features is shown in Figure 4. Feature tags are actually class names, and features have all details of their dimensions stored. Different XML elements are used when feature and process data are used in system context (e.g. cell formation, scheduling, FMS control) since for system context only high level routing data that contain process sequences, machines, tools and process times are necessary.

4.2 XML DATA WRITING AND PARSING

An object-oriented data generation framework was constructed for the purpose of generating XML data for an FMS models based on the process plan representation model. The different entities whose XML data needs to be generated for the FMS model include _MfgPartModel_, _Machine_, _PartActivity_, _MachineActivity_ and _ToolDirectionActivity_, _MfgFeature_ and _MfgProcess_. All the Java classes built for the process plan representation model are capable of generating self-describing XML data. We developed two alternative ways for achieving this
Integration of Product Design, Process Planning and FMS Control using Neutral Data Representation

goal: a) each class implements methods for XML data generation, or b) XML data generation is delegated to special XML writer classes.

![XML data for features](image)

The first approach was used for `Machine` class by declaring and implementing method `writeXML()`. The `Machine` class holds data about the machine like name, number of units, usage frequency, SICGE code etc. It is this state data for each machine that needs to be saved into XML.

![XML processing for classes](image)

Alternative approach was to implement XML writer class for each model class (see Figure 5). The XML writers write out XML data at a given state of manufacturing process planning. The `PartWriter` writes part material and part activities. It calls `FeatureWriter` to write the feature data. The feature writers also work in the same manner. The `FeatureWriter` writes the dimensional details like radius, depth, axis and axis-point about the feature. It gives a call to the `PropertyTableWriter` which writes the tolerance details like surface finish, roundness, true position and straightness. It also calls `ProcessWriter` that writes data for alternative processes like, class name (type), process name, cutting parameters, machine name, process time and cost. This chain of XML data writing is shown in Figure 5a.

It should be noted that during the writing process writers verify the state of the process planning and write only data that is available. For example, at the end of feature mapping only feature details are written, and the writer does not write anything about processes or tolerances.

XML data generated from FMS model holds all the information necessary to reconstruct the Java class instance, including data about all the parts in the FMS model, the machines in the FMS model, the routing information for all the parts etc. XML parsers are used to reconstruct instances of Java classes from the XML data. There are two kinds of parsers available for XML data parsing [7]: SAX parser (Simple API for XML parsing) and DOM parser.
A SAX parser, unlike the DOM parser, does not maintain a default model for the parsed data. If a listener can be set on the SAX parser to listen to parsing of specific tags, then whenever the corresponding element is parsed, the listener can construct a Java instance of the class corresponding to the element. The SAX parser used in FMS model is an instance of the Java class \textit{(javax.xml.parsers.SAXParser)}. This parser is used to construct FMS models from the saved XML files. The listener for this parser is called as content handler and is an instance of the Java class \textit{(org.xml.sax.helpers.DefaultHandler)}. Every object (or class) in the FMS model is responsible for defining and implementing a content handler that can reconstruct instances of the Java class from the XML element parsed. For example, for the 	extit{Machine} class, the content handler is defined as an inner class and an instance of the content handler can be obtained by calling a method called \textit{getSAXHandler()}. XML data can then be parsed with this content handler.

The content handler provides three methods that are used to reconstruct the wrapper class instance: \textit{startElement()}, \textit{endElement()}, and \textit{characters()}. Similarly to XML writing procedure, XML parsing process is also delegated to individual classes. This is accomplished by holding a stack of content handlers responsible for parsing the complete FMS model representation. This is illustrated in Figure 5b. For example, \textit{MfgPartModel} object is created from XML file by \textit{PartHandler}, which understand data such part name, material, batch size, etc. When this handler finds \textit{<featuresList>} tag it delegates feature reconstruction to another handler, \textit{FeatureHandler}, which restores feature dimensions for the current feature. The \textit{FeatureHandler} is replaced by \textit{ToleranceHandler} and \textit{ProcessHandler} when it finds corresponding tags. Those two handlers restore tolerance and process data respectively from XML file.

5. INTEGRATION OF FEATURE RECOGNITION, PROCESS PLANNING AND FMS SIMULATION

Utilization of XML-based process plan data model will be demonstrated in distributed integration of process planning and FMS Control. Scenario is the following. For a group of designed parts it is necessary to perform process planning and select the process plan from several alternatives in order to design FMS control module, which will be studied by simulation. The modules integrated into this scenario are shown in Figure 6. Feature Mapping module is described in \cite{7}, while rule based process planner, which is part of IMPlanner system is described in \cite{13}. FMS Control simulation model is briefly described in this section.

FMS simulation model represents FMS control procedures. It is developed as a discrete event model in Arena \textit{Error! Reference source not found.}. It is capable of applying several control policies for selection of alternative routings, and several dispatching rules to order processing of parts on machines. Control policies include best plan, random plan, and dynamic model with feature focus which selects alternative for each manufacturing feature. The last model, which we call feature focused dynamic model requires integrated approach in which all three modules (feature mapping, process planning and FMS controller) participate.

In a feature focused dynamic simulation model, appropriate machining processes are generated on demand by running rule based system. This rule based system not only selects capable processes for the feature but also selects alternative machines available to make that feature with tool and processing time information. Alternatives for every part are sent to the FMS controller which is responsible for collecting the data about the FMS status and then selects the best process plan for this status to be used in the next manufacturing period. This module will provide required information to simulation model when prompted.

In this novel approach, at every predefined time interval, the simulation model will collect and send current machine utilization or queue size for every machine to the FMS controller. Based on the information fed into the FMS controller module, this module will select the best machine in terms of lowest machine utilization or queue size out of alternatives to make each feature. After selecting the best machines, process plan in simulation model is updated with the best machines for part routing data. New entities arriving at this point of time uses new updated process plan. This selection of process plan continues at same predefined time interval.

Each entity runs through simulated FMS following given process plan. Five dispatching rules are applied for prioritizing between parts for processing at each machine. These dispatching rules are i) FIFO, ii) SPT, iii) SIPT, iv) LIPT, and v) Dissimilarity Maximization/Minimization. After the completion of all the operation for each part, the part leaves the system and its performance measure are recorded. The performance measures considered are machine utilization, average flow time, and work in progress.
6. **CASE STUDY**

This scenario has been tested on several mechanical parts from NIST design repository [8]. Six different parts were chosen as shown in Figure 7. For each part CAD model was modified to feature based model to accommodate our approach to CAD/CAPP integration. Each model has been separately run by feature mapping application (number of features per part varies from 7 to 18). The results for all models are saved into XML files with the full feature details. Process selection is done by rule-based system which generates alternative processing for each feature and selects four different alternatives for each part. The results of those two modules are illustrated in [10]. Simulation model, as described above has been run for three scenarios: best process plan, random process plan and dynamic process plan. Preliminary results of running model for best and random process plan show a potential improvement of 20% reduction in production cycle. The feature based dynamic model has been implemented, and simulation experiments are expected to provide similar improvement. For experimentation, different scenarios were generated by combining different routing selection policies and dispatching rules. After running each scenario the performance measures of each scenario were recorded, statistically analyzed and compared to observe the effect of each combination on the FMS model.

7. **CONCLUSIONS**

This paper has demonstrated the flexible use of an XML data representation in process planning modeling and integration. The XML schema has been developed for validation of XML process plan files. Flexible approach for writing the XML data from an FMS model and parsing data back has been explained. Benefits of this approach were demonstrated in few small examples. Scenario for the use of proposed XML based model in integration of feature mapping, process planning, and FMS control using a simulation model is demonstrated on simple example.
<table>
<thead>
<tr>
<th>Name</th>
<th>AES94</th>
<th>Plate</th>
<th>Netex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td><img src="image1.png" alt="Model" /></td>
<td><img src="image2.png" alt="Model" /></td>
<td><img src="image3.png" alt="Model" /></td>
</tr>
<tr>
<td>Number of features</td>
<td>10</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>USC</th>
<th>Slider</th>
<th>Bracket</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td><img src="image4.png" alt="Model" /></td>
<td><img src="image5.png" alt="Model" /></td>
<td><img src="image6.png" alt="Model" /></td>
</tr>
<tr>
<td>Number of features</td>
<td>6</td>
<td>18</td>
<td>11</td>
</tr>
</tbody>
</table>

Figure 7. Part designs used in case study

REFERENCES


Application of the Design of Experiments to Evaluate Production Efficiency in Lean Systems

Steve Davies*, Tim Coole, and David Osypiw
Faculty of Creativity and Culture
Buckinghamshire New University
High Wycombe, Buckinghamshire, HP11 2JZ, United Kingdom

ABSTRACT

Lean Manufacturing Systems strive to achieve customer value through the elimination of waste by synchronising production output to the pull of customer demand within a culture of continuous improvement. The challenge for manufacturers that utilise resources that produce batches of products is to maximise the productive efficiency of the resource. Maximising the productive efficiency of the resource enables minimised production batch sizes that do not jeopardise customer supply, optimise the use of factory and storage space and packaging requirements. Manufacturing processes are subject to variation, both in respect to the inputs to the process and to the process output itself. Consequently, there are inherent difficulties in determining how to maximise production efficiency due to the nature of the variation within the system not being fully understood. This paper examines the application of the Design of Experiments methodology to understand how inputs into a manufacturing process both singularly and through interaction influence the performance metric of the process. The paper considers a case study application in a car body panel pressing facility where the factors that determine the performance metric of a production press are analysed through the creation of the appropriate factorial design. The analysis contributes to an evolving body of knowledge with respect to the panel production process and to aid the identification of lean batch sizes that do not compromise performance.

1. INTRODUCTION

The purpose of a lean system whether in manufacturing, service or administration is to provide customer value through the minimum application of resources. Such systems are further characterised by synchronising the work flow through the system to the pull rate of the customer by creating value added activities within a culture of continuous improvement [1].

Matching production output to the pull rate of the customer is a particular challenge to manufacturers who are constrained to produce batch quantities of products. Generally, the production rate of the manufacturing resource is far in excess of the pull rate of the customer. Economic use of the resource is achieved by producing multiple products that will consume the available resource capacity. In a lean environment, a further challenge for batch manufacturers is minimizing the size of the batch runs while not compromising the performance of their production resources. It is necessary therefore to understand the conditions where the performance levels operate at optimised conditions.

Vehicle body manufactures are one such set of producers that are constrained to manufacturing batch lot sizes of various body panel types relative to the volume build requirements of their vehicles. Specialist body panel suppliers are further constrained by the need to supply panels to several vehicle manufactures each with a different build volume requirement and assembly rate. The panel supplier is therefore dealing with multiple pull rates across common production facilities. To meet these constraints, the panel manufacturer must ensure that the production performance rates are such that the demand for each body panel is satisfied. Demonstrated production rates are variable and are influenced by a number of factors including the batch size, the set up time and the speed the production press is set to run the panel. However, while each of these factors has planned or target quantities, during the actual production run, they can vary. Moreover, during the production run, unplanned downtime will sporadically occur that will compromise the performance rate.

* Corresponding author: Tel.: (44) 1793 871101; E-mail: rdavie01@bucks.ac.uk
Given that the performance rate of a production press is influenced by variable factors, this paper examines the application of statistical methods within the discipline of the ‘Design of Experiments’ to understand how such factors both singularly and through interaction influence the performance rate. The purpose of the analysis is to identify the conditions where minimum batch run lengths can be identified that do not compromise the performance rate.

The analysis is conducted through a case study of the production performance of a high volume vehicle body panel manufacturer. The case study environment is characterised by attaining a high degree of quality performance with respect to both body panel dimensional and cosmetic attributes. However, the environment is further characterised by inconsistency in maintaining target set up times and batch run lengths and with respect to the underlying net running rate of a production press. These inconsistencies combine to influence the overall production rate.

Section 2 introduces the discipline of the Design of Experiments (DOE) method and in particular the Analysis of Variance (ANOVA). To aid the manufactures understanding of the DOE method, the reason why the ANOVA procedure can identify if a factor or factor interaction influences the measured response of interest is discussed.

Section 3 introduces the case study and presents the analysis from the experimental design that quantifies the effect of the factors and factor interaction on the performance rate.

Section 4 provides concluding remarks and assesses the potential of the DOE approach and identifies further applications of experimental design within the case study environment.

2. THE DESIGN OF EXPERIMENTS

The discipline of the ‘Design of Experiments’ is a proven methodology that has been successfully applied across a number of diverse disciplines including manufacturing, engineering, biology, the social sciences and psychology that quantifies the relationship between measurable process inputs and their influence on a measured process response. Within the last twenty years within the ‘quality improvement’ community, DOE has become accepted as one of the most powerful improvement methods to reduce process variation, enhance process effectiveness and process capability [2]. Succinctly, DOE provides a structured methodology that provides a procedure to observe how an output from a process is affected by specific changes that are made to the settings of the process inputs [3].

The discipline evolved from the work of the eminent British statistician Sir Ronald Fisher during the 1920’s while conducting research into increasing crop yield at the Rothamsted Agricultural Experimental Research Station in Hertfordshire, England. Prior to Ronald Fisher arriving at Rothamsted, the experimentation to determine the effect of different types of fertilizers on crop yield was through applying a single type of fertilizer to a field for a single year and measuring the crop yield. The following year another type of fertilizer was applied to a field and a different fertilizer the following year and so on. Over several years, the research centre would compare the fertilizer contribution to crop yield taking into account attributes such as temperature and rainfall and sundry environmental factors. Upon arriving at Rothamsted, Fisher concluded that despite the rigorous and logical approach taken by the previous experimenters, it was not possible to formulate any worthwhile conclusions from the analysis of the crop yield data. It was not possible to determine if observed differences in the crop yield over the years was due to the effect of the different fertilizers or to environmental factors such as rainfall, land drainage or sunlight. Fisher introduced an alternative approach to experimentation by comparing the effects of all the fertilizers simultaneously through segregating different plots of land into blocks and randomly assigning the fertilizers to the blocks and through the replication of experiments to supply an estimate of the error. The logic of the randomisation implied that any differences in crop yield would be attributable to the choice of fertilizer and not to any one plot or block receiving more rainfall, sunlight or drainage [4].

Fisher recognised that the source of the variation within a set of observed data was either random or attributable to the influence of the factor effects (singularly or through interaction). So, while planning and implementing his new method of experimentation, Fisher also developed the statistical techniques necessary, in particular the methods of the ANOVA, to identify the source of the variation within the data gathered through the experimentation. In 1925, Fisher published his ‘Statistical Methods for Research Workers’ in which he describes the statistical methods that form the foundation for experimental design. Fisher developed his ideas further in 1926 with the publication of ‘The Arrangement of Field Experiments’ and in 1935 with the ‘Design of Experiments’ where the ideas of randomization, blocking and replication are further explored [5].
2.1. DESIGN STRUCTURE

Experimental design derived from the work of Fisher is based on three fundamental techniques of randomisation, replication and blocking [6]. Randomisation prevents both systematic and personal bias influencing the values of the measured response variable. Replication involves the repetition of an experimental run to obtain independent observations of each factor combination. Replication estimates the experimental error and is a basic unit of measurement for determining if the observed differences in the data are statistically different [7]. Blocking improves the precision of the comparisons between factors by reducing or eliminating the variability due to nuisance factors or systematic variation. Blocking partitions the observations into a set of groups called blocks such that the observations are collected under homogeneous experimental conditions.

The terminology used within an experimental design is listed in Table 1 and is consistent across all types of design.

Table 1: Design of Experiments Terminology

<table>
<thead>
<tr>
<th>Design Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured Response Variable</td>
<td>The measured process output variable that is the subject of the experiment.</td>
</tr>
<tr>
<td>Factor</td>
<td>An input variable to the process. A factor may be controllable in that it can be fixed for the experiment or uncontrollable in that it cannot be precisely fixed for the experiment.</td>
</tr>
<tr>
<td>Treatment Level</td>
<td>The levels that a factor can be applied to a process are referred to as a treatment. Treatments may be quantitative (pressure, temperature) or qualitative (material types, gender).</td>
</tr>
<tr>
<td>Main Effect</td>
<td>The effect on the measured response variable that is attributed to changing the treatment level of a single factor.</td>
</tr>
<tr>
<td>Interaction Effect</td>
<td>The effect on the measured response variable that is attributed to two or more factor interactions.</td>
</tr>
</tbody>
</table>

Fisher’s approach to quantifying the variation of a system was to propose that the system had an underlying constant mean response output such that any random observation of the output deviated from the mean due to the influence of the treatment levels of a factor and random error. Formally, if \( n \) random observations are made of a single factor applied at \( a \) treatment levels, an observed response \( y_{ij} \) is specified as a deviation from an overall response grand mean \( \mu \) such that

\[
y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad i = 1,2,\ldots,a \quad j = 1,2,\ldots,n
\]

where \( \epsilon_{ij} \sim N(0,\sigma^2) \) independently of one another and \( \sigma^2 \) is the common variance within all treatments.

In Equation 1, \( \mu \) is the overall mean of the total observations and \( \tau_i \) is a parameter unique to the \( i \)th treatment called the ‘Effects Model’. Because Equation 1 investigates the effect of one factor, the equation is modelled by the One-Way ANOVA method to determine if the treatment means are equal. The One-Way ANOVA model partitions the total variation of the observed responses between the variation due to the treatment effects and the variation due to random error. If the treatment means are not statistically significantly different then the overall variation is due to the random errors within each treatment. Formally, the total variation is defined as the total sum of the squared deviations from the grand mean and is denoted by \( SS_T \) (for Sum of Squares Total). The treatment partition is denoted by \( SS_{\text{Treatments}} \) and the error partition is denoted by \( SS_E \) to give

\[
SS_T = SS_{\text{Treatments}} + SS_E
\]
The variances of each of the terms in Equation 2 are obtained by dividing each term by their respective degrees of freedom. In total there are \( an = N \) random observations yielding \( N-1 \) degrees of freedom for \( SS_T \). There are \( a \) treatment levels of the factor that will yield \( a - 1 \) degrees of freedom for \( SS_{Treatments} \). And with respect to the random error, due to the \( a \) treatment levels of \( n \) observations, \( SS_E \) will have \( a(n - 1) = N - a \) degrees of freedom. The variance components on the right hand side of Equation 2 are commonly referred to as ‘Mean Squares’ or simply \( MS_{Treatments} \) and \( MS_E \) respectively. Montgomery [7] establishes that the \( MS_E \) is an unbiased estimator of \( \sigma^2 \), and under the condition that the treatment means are equal then the \( MS_{Treatments} \) is also an unbiased estimator of \( \sigma^2 \). Consequently, if the null hypothesis of no differences in treatment means is true, then

\[
F_0 = \frac{MS_{Treatments}}{MS_E}
\]

(3)

follows an \( F \) distribution with \( a - 1 \) and \( N - a \) degrees of freedom. Equation 3 is the test statistic for the null hypothesis of the equality of treatment means where

\[
H_0 : \mu_1 = \mu_2 = \ldots = \mu_a \\
H_1 : \mu_i \neq \mu_j \text{ for at least one pair } (i, j)
\]

(4)

For a given significance level \( \alpha \), the null hypothesis \( H_0 \) is rejected if

\[
F_0 > F_{a-1,N-a,\alpha}
\]

(5)

The test statistic \( F_0 \) (Equation 3) and the decision criteria specified in the Inequality (5) form the F Test (in honour of Ronald Fisher) and provides a method for determining the equality of variances between different populations. The F Test forms the hub of the DOE analysis as the test confirms if a factor applied at different treatment levels to a process influences process response. The F Test applies to both Randomised Block Designs and Factorial Designs. In a randomised block design a parameter \( \beta \) is introduced to quantify the variation due to the blocking factor giving an effects model

\[
y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}
\]

(6)

The partitioning of the total sum of squares \( SS_T \) yields

\[
SS_T = SS_{Treatments} + SS_{Blocks} + SS_E
\]

(7)

For a factorial design, the partitioning of the total variation is dependent on the number of factors used in the design. Given a design that includes three factors (specified as a 2 \( ^3 \) Factorial design) where the main effects are modelled by the parameters \( \tau_i, \beta_j, \) and \( \gamma_k \), the effects model is

\[
y_{ij} = \mu + \tau_i + \beta_j + \gamma_k + \tau \beta_{ij} + \tau \gamma_{ik} + \beta \gamma_{jk} + \tau \beta \gamma_{ijk} + \epsilon_{ijk}
\]

(8)
The algebraic decomposition of the total sum of squares for each of effects models to partition the overall variation is well defined in the DOE literature, including Montgomery [7] and Hicks and Taylor [8]. In each of the models, significant effects are quantified through carrying out the appropriate F test between each of the factors, their interactions and the random error to determine the equality of their respective variances.

3. Case Study: Application of Design of Experiments

The basic manufacturing method to produce a body panel consists of cutting flat blanks from coil steel that will roughly match the contour of the panel followed by a series of pressing operations to produce the finished panel. The first of the pressing operations, feeds the flat blank through a ‘draw die’ to form the basic three dimensional shape of the panel. Additional pressing operations add apertures (for windows and doors), piercings to allow access for fasteners, wiring looms and trim parts. Final forming and trimming operations complete the finished panel. Generally, there can be up to six pressing operations to complete the finished panel, though modern press tooling can produce very complex body panels within four pressing operations. The press tools are held within a power press of which there are two basic kinds, a Tandem press line or a Transfer press. A Tandem press line consists of a series of separate presses that either use automation or manual operators to transfer the panel between pressing operations. A transfer press is a self contained unit in which all the press tools are located and an automation system transfers the panel between the tooling operations.

Vehicle body panel production poses a challenging environment for manufacturers. There is the challenge to meet and continually improve on rigorous customer quality standards with respect to both cosmetic and dimensional attributes of the panels. And of equal importance, is the challenge to satisfy demanding customer delivery schedules for the panels. Whether the body panels are produced in a pressing facility that is a subdivision of the vehicle manufacturer or at a specialist body panel supplier, the pressing facility will produce a variety of panels for a range of vehicle model types. Additionally, the specialist body panel supplier will supply body panels to several vehicle manufacturers.

Panel manufactures by necessity produce in batches. However, batch production compromises the lean principle of flow where the goal is to synchronise manufacturing output to the demand or pull rate of the customer. Over time, the customer pull rate on any given press line will vary due to the vehicle model mix requirements from each vehicle producer. Given the variation in customer ‘pull rates’, the panel manufacturer must ensure that the output rate of their production resources can satisfy any imposed combination of demand from any of their customers. Consequently, a key question a panel manufacturer attempts to answer is under what running conditions will optimised performance be achieved? Certainly, the continued maintenance of the press lines and press tools along with ongoing set up time reduction programs will provide the foundation for ensuring optimised running performance. However, the panel running conditions are variable. Planned batch sizes vary by panel type and model mix, actual set up times vary, and the underlying speed of the press varies throughout the production run due to the specific running conditions experienced during the run. How this combination of variation affects the performance rate of a press line will also contribute to understanding how to achieve an optimised production rate.

This case study considers a panel pressing facility consisting of some fifteen transfer and tandem press lines that are capable of producing a complete range of body panels for a diverse model range for several vehicle manufacturers. Each press line is linked to a production monitoring system that records the production achievement, set up duration, downtime and calculates performance rates. Two key production performance rates are calculated by the monitoring system for each panel production run. The first is defined as the Gross Shots per Operating Hour (GSPH) and is a measure of the number of panels produced during the run divided by the total time to complete the run. The second rate is defined as the Shots per Operating Hour (SPOH) and is a measure of the panels produced divided by the time the press was running (so excludes the set up time and down time recorded during the run). The GSPH rate is the key performance measure for a press line. An average GSPH rate is calculated for each panel (usually a rolling average of the previous eight production runs) and aggregated across the requirement for the press line to determine the capacity requirement for the line. The capacity requirement is used to establish shift patterns, manpower requirements and the sourcing of work to press lines. Other business activities are also based on the capacity statement including maintenance and financial planning. Maximising the GSPH rate is a key requirement to minimizing the capacity requirement for the pressing facility to meet customer demand.

3.1. Design Model
The measured response of interest is the GSPH performance rate of a production press. In particular the manufacturer is interested in maintaining an optimised rate while attempting to minimise the planned batch size. The manufacturer is aware that other variable factors contribute to the level GSPH rate and in particular the net running rate of the press specified by the SPOH and the set up time from changing from one tool set to another. Each of these factors cannot be set in advance as they are also variable outputs of the panel pressing process. The design is therefore based on observed data taken from the production performance database. The chosen factor levels are in fact mean values of the factor for some interval around the mean value. The chosen factor levels are set up times of 10 and 20 minutes, for the SPOH, 475 and 575 strokes per hour and for the achieved batch run lengths 2000 and 3000 panels. The observed data presented in Table 2 is returned from production records for a single production press line.

Table 2 illustrates the $2^3$ design that is created from taking four random selections of the GSPH rates at each level from the data observations that satisfied the factor combinations.

<table>
<thead>
<tr>
<th>Run Length</th>
<th>Run Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set Up</td>
<td>SPOH 2000</td>
</tr>
<tr>
<td>10</td>
<td>475</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>575</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2. MODEL ANALYSIS

The design defined in Table 2 is input into the DOE function of the Minitab Statistical Analysis Package with a chosen significance level for the test statistic set at $\alpha = 0.05$. The Minitab package in common with other statistical packages carries out the partitioning calculations and calculates the test statistic for each factor and factor interaction. The Minitab output is reproduced in Table 3. Though Minitab calculates the values of the test statistic $F_0$, the critical $F_{\alpha-1,N-\alpha}$ value is not provided. Minitab relies exclusively on specifying the ‘P’ value to determine if a factor or factor interaction has an effect on the measured response variable. Under the assumption that the null hypothesis is true, the P value returns the probability of obtaining values of the test statistic that are equal or greater in magnitude than the observed test statistic. Therefore, if a P value is returned that is a smaller value than the significance level for the hypothesis test, then evidence exists to reject the null hypothesis.

In the Analysis table, the SPOH rate is the only factor to have a P value less than the specified significance level of 0.05 implying that the null hypothesis is rejected with respect to the effect of the SPOH rate. The other main effect factors, the Run Length and Set Up and the 2 and 3 way interactions have no bearing on the value of the GSPH rate. This conclusion is illustrated graphically in the ‘Normal Probability Plot of Standardized Effects’ plot and the ‘Pareto Chart of Standardized Effects’ chart obtained from Minitab. In each chart the SPOH is the only factor to show a significant effect. In the Normal Probability Plot, significant factors are highlighted in red and in general will be further from the normal line compared to the non-significant factors. In the Pareto Chart, the bar for a significant factor will extend past the red line. The position of the red line is drawn at the $(1 - \alpha/2)$ quantile of a t-distribution with degrees of freedom equal to the degrees of freedom for the error term [9]. Since the error term has 24 degrees of freedom, at the 0.05 significance level, a value of 2.064 is returned from t distribution tables.

The SPOH rate defines the continuous average running speed of the press. It is also an indicator of how well the press and press tooling are performing during a production run. If the press and tooling are performing well, the running speed of the press is increased (often slightly over the target press speed). If there are difficulties during the run, the speed of the press will be lowered to enable the production of quality panels. The analysis confirms the need
to focus continuous improvement activities to ensure that the press and associated tooling will consistently perform toward the upper bound of the press speed capability. Critically, the analysis confirms that achieved batch run length does not contribute to the level of the GSPH rate. While the decision to choose a specific batch size for a panel production run will take into consideration such criteria as the demand profile, container and storage requirements, the analysis provides confidence to a manufacturer to choose a lower batch size without compromising performance efficiency. Though the set up time is not considered a significant factor, the manufacturer understands that the level of the GSPH rate is dominated by the level of the SPOH rate, and the analysis does not mitigate the need to continually engage in set up reduction programmes. The absence of factor interaction indicates to the manufacturer that each factor contribution can be assessed independently of each other with respect to improving performance.

Table 3: Minitab Analysis of 2³ Factorial Design to Assess Effect on GSPH Rate

Factorial Fit: Rate versus Set Up, SPOH, Run Length

Estimated Effects and Coefficients for Rate (coded units)

<table>
<thead>
<tr>
<th>Term</th>
<th>Effect</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>378.63</td>
<td>10.45</td>
<td>36.23</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Set Up</td>
<td>7.23</td>
<td>3.62</td>
<td>10.45</td>
<td>0.35</td>
<td>0.732</td>
</tr>
<tr>
<td>SPOH</td>
<td>80.38</td>
<td>40.19</td>
<td>10.45</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Run Length</td>
<td>-12.23</td>
<td>-6.12</td>
<td>10.45</td>
<td>-0.59</td>
<td>0.564</td>
</tr>
<tr>
<td>Set Up*SPOH</td>
<td>-21.37</td>
<td>-10.69</td>
<td>10.45</td>
<td>-1.02</td>
<td>0.317</td>
</tr>
<tr>
<td>Set Up*Run Length</td>
<td>5.43</td>
<td>2.71</td>
<td>10.45</td>
<td>0.26</td>
<td>0.797</td>
</tr>
<tr>
<td>SPOH*Run Length</td>
<td>-15.85</td>
<td>-7.93</td>
<td>10.45</td>
<td>-0.76</td>
<td>0.456</td>
</tr>
<tr>
<td>Set Up<em>SPOH</em>Run Length</td>
<td>9.56</td>
<td>4.78</td>
<td>10.45</td>
<td>0.46</td>
<td>0.652</td>
</tr>
</tbody>
</table>

S = 59.1105  R-Sq = 41.68%  R-Sq(adj) = 24.67%

Analysis of Variance for Rate (coded units)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td>3</td>
<td>53306</td>
<td>53305.9</td>
<td>17768.6</td>
<td>5.09</td>
<td>0.007</td>
</tr>
<tr>
<td>2-Way Interactions</td>
<td>3</td>
<td>5900</td>
<td>5899.6</td>
<td>1966.5</td>
<td>0.56</td>
<td>0.645</td>
</tr>
<tr>
<td>3-Way Interactions</td>
<td>1</td>
<td>731</td>
<td>730.5</td>
<td>730.5</td>
<td>0.21</td>
<td>0.652</td>
</tr>
<tr>
<td>Residual Error</td>
<td>24</td>
<td>83857</td>
<td>83857.4</td>
<td>3494.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pure Error</td>
<td>24</td>
<td>83857</td>
<td>83857.4</td>
<td>3494.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>143793</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4 CONCLUSIONS

Vehicle body panel manufacturing is a complex process. The manufacturer understands that to determine optimised production conditions, no one single analytical method will provide the full answer regardless of the methods proven effectiveness in other manufacturing and non-manufacturing applications. Rather, an analytical method has the potential to contribute to an ongoing body of knowledge with respect on how to manage the pressing operations. Within the specific application of DOE in this case study this is certainly the case. The choice of batch size for example is dependant on several factors but primarily the volume of demand and container and warehouse storage requirements. The DOE analysis has provided confidence to the manufacturer that they can choose a lower bound on a batch size without compromising performance. Moreover, the analysis has directed focus onto improving tooling and press operating performance to increase the SPOH rate that in turn will increase the overall GSPH rate. While the set up time is not considered a contributing factor with respect to the GSPH rate, the manufacturer understands that the set up effect is negligible when compared to the effect of the SPOH rate and that it does not mitigate the need to continue to engage in set up time reduction programmes.

This particular experiment relied on observational data as the input factors could not be fixed. This led to practical problems when setting up the experiment. Firstly a factors value had to be set and random observations taken from with a restricted interval around the factor value. As each factor was taken into account a reduced number of observations intersected each of the factor intervals. Possibly, if a further factor was introduced to the experiment it is likely that there would not have been enough intersecting observations to carry out a valid experiment. What made this experiment possible was the existence of the production performance database. In a similar manufacturing environment, the lack of a similar database would prevent the carrying out of this type of experiment.

However the absence of a performance database does not preclude the application of experimental design to assess performance characteristics of the panel manufacturing process with controlled factors. For example performance characteristics can be compared between press lines of the same type (press size and tonnage). The individual press lines would form the blocks for a Randomised Block Design such that the variation due to the press lines will become apparent due to the blocking effect in the output of the experiment. The measured response of the experiment need not be constrained to the performance rate, but can apply to any measured response from the panel manufacturing process, for example set up time duration or downtime periods. Further experiments could determine if production characteristics differ due to tooling type or panel type (deep draw, shallow draw, outer skin panel, internal panel). Such experiments would need to be complemented by a comparison method such as the Tukey method for all pair wise comparisons or the Schef’s method for all comparisons to quantify which factor levels are significantly different from one another [10]. Generally the comparison methods are integral to the DOE and ANOVA functionality within available statistical analysis packages.

REFERENCES

Applying Lean Principles to Production Scheduling

Mustafa Ramzi Salman\textsuperscript{1}, Roman van der Krogt\textsuperscript{2}, James Little\textsuperscript{2} and John Geraghty\textsuperscript{1}

\textsuperscript{1}Enterpris Process Research Centre, Mechanical and Manufacturing Engineering, Dublin City University, Dublin, Ireland
\textsuperscript{2}Cork Constraint Computation Centre, Department of Computer Science, University College Cork, Cork, Ireland

Abstract

When exploring the Lean issue, literature appears to have countless definitions to describe that individual topic, causing misinterpretations between academics and professionals alike. The mainstream Lean implementation ventures are merely a compilation of tools and methods that are forced down the organisational hierarchy from higher tiers, due to other organisations publicising lean in improving their enterprise’s effectiveness and efficiency. However, lean implementations have often undesirable outcomes. Firstly, organisations either do not achieve the levels of success they initially yearned. This leads to management either deserting their lean efforts or seeking aimlessly for a new buzz word to adopt. Secondly, an enterprise may find it difficult to sustain the lean improvements, and thus rushing to conclude lean as incompatible with their industry, or unsuited for their organisational structure. These shortcomings can be easily avoided by recognising lean not only as a collection of tools but also a philosophy. The Lean philosophy needs to be applied companywide, with both management and employees onboard, to tackle business issues as they arise with a united way of thinking. In turn, this means that when opportunities or problems occur, and traditional lean tools fail short, the shared way of thinking by the organisation can be used to derive and develop new tailored solutions to address the issue directly. This paper aims to illustrate the possibilities of lean tools being applied in complex production environment, and introduce two case studies supporting the findings. In particular, we argue that the full potential of planning and scheduling in a manufacturing setting can only benefit from the utilisation of lean principles and the support of quantitative modelling.

1. Introduction

While there are many methodologies to increase the efficiency of production operations, it is often the case that there is a need to apply a combination of tools to achieve the desired objectives. It is with that in mind that this paper intends to enlighten the merits of applying lean principles to production scheduling.

The argument presented here consists of two parts; firstly, it is concerned with the adoption of Lean in an appropriate manner and not attempting to imitate other organisations or tracing their footsteps, as this has shown to be a common explanation for companies failing to realise or achieve Lean success. Secondly, Lean can venture beyond its automobile origin, known to be Toyota in Japan, to benefit other industries and market sectors. Here the concept of applying Lean to production scheduling is investigated in more detail. Lean tools that can be adapted to scheduling are explained and analysed to help highlight the agenda. This is further emphasised by presenting two case studies, where Lean tools and techniques are implemented to enhance the scheduling effectiveness. The first case study shows the impact ‘-leaning’ (particularly, buffer reduction) has on the scheduling process and how quantitative modelling can be used to examine this. The second case study investigates the support of day-to-day operations in a Lean-centric cellular manufacturing environment.

The remainder of the text is structured as follows. The next section briefly summarises the Lean philosophy, followed in Section 3 by a review of scheduling, accompanied by Lean-tools of interest to scheduling. Then, in Section 4, the two case studies are presented. The paper closes with a discussion and conclusion.

* Corresponding author: Tel.: (+353) 866055094; E-mail: mustafa.salman2@mail.dcu.ie
2. LEAN PRODUCTION

The term Lean was first used by Krafick [1] in 1988, a principle researcher in the International Motor Vehicle Program (IMVP) at Massachussets Institute of Technology (MIT), to describe what today is known as the Lean Manufacturing or Lean Production paradigm [2]. A major output of the IMVP research efforts was the publication of the book ‘The Machine that Changed the World: The story of lean production’ [3]. This book documented the evolution of the automotive industry from Craft Production, through Mass Production to ultimately Lean Production. The book argues that the Lean paradigm follows in the footsteps of its peers, but only by combining their advantages, while trying to evade the high costs of craft producers, and the rigidity of mass producers. Lean is a philosophy that seeks to improve activities by exposing and eliminating all types of waste from the system, may it be a value stream, a manufacturing process, or even a routine job. It makes use of its tools to strive for zero inventories, zero downtimes, zero defects, and zero delays in the production process. Waste can be described as the opposite side of value on a Lean coin. Value is all the aspects of a product that a customer is willing to spend his/her money on [4]. Lean emphasises the need for an organisation to develop a culture of continuous improvement in quality, cost, delivery and design [5]. Lean has been described as a quest for “brilliant process management” [6]. Lean asks an organisation to consider all activities as a series of processes that can be optimised and aligned to continuous improvement programmes.

2.1. THE TOYOTA PRODUCTION SYSTEM, JUST-IN-TIME AND LEAN PRODUCTION

A point of confusion for both academic researchers and industrial practitioners is the relationship between the Toyota Production System, Just-In-Time and Lean Production. Sugimori et al. [7] describe TPS as comprising two main concepts; ‘Cost reduction through the elimination of waste’, and ‘Full utilization of worker’s capabilities’. Cost reduction is primarily achieved through the use of Just-In-Time production (Kanban, one-piece flow and levelling) and Jidoka; operators have authorisation to stop the line to prevent defective work being transferred to the next station [8, 9, 10]. TPS achieves full utilisation of worker’s capabilities by promoting respect for individuals through minimising employee movements, emphasising employee safety, valuing and encouraging employee involvement, and increasing employee responsibilities [11].

Some authors argue that Lean Production is an improvement or descendant of TPS and JIT [12], while others argue that it is merely a Westernised version of TPS and JIT [13]. Papadopoulou & Özbayrak [2] argue that evidence from the literature suggests that Lean Production and JIT should not be considered as systems, rather as philosophies that have a broad scope and are continually under development and “have neither exact nor sole definition nor obvious boarders”. However, when they compared the American Production and Inventory Control Society (APICS) Dictionary definitions of JIT and Lean they concluded that the similarities in the definitions far outweigh the differences, thereby, giving support to the idea that Lean is a descendant of the JIT philosophy.

Lander and Liker [11] identified that many organisations struggle with understanding how Lean Production principles can be applied to them. When they look at Toyota, they see highly repetitive, standardised manufacturing processes and have great difficulty in seeing how they can transfer this to their organisations, which may build-to-order, involve highly engineered products or they may be service providers. Lander and Liker [11] argue that these companies are approaching Lean Production from the wrong perspective. What is required to transfer the benefits of Lean to environments where they have yet to be applied is to adopt the same principles and thought processes that Toyota used to develop TPS, rather than direct implementation of a set of tools and techniques from a toolkit. Viewed in this light TPS is not a production philosophy, but rather an example of the application of JIT and Lean philosophies to one organisation in the automotive sector.

Lean offers opportunities for many organisations, to increase productivity, reduce waiting times, lower costs and improve services, if it can be successfully applied in their business sector. The literature on Lean is rich with case studies on the successful application of Lean in a variety of sectors worldwide (e.g. Liker, 1998; Drickhamer, 2000; Siekman, 2000; Zimmer, 2000; Lewis, 2001; Prizinsky, 2001; Strozniaik, 2001; Bateman, 2002; Rea, 2002; Teresko, 2002; Trombly, 2002; Parker, 2003).

2.2. THE ELEMENTS OF LEAN

One of the trademark features of Lean is its reliance on visual and simple mechanical aids to improve manufacturing effectiveness. Most fundamentals of Lean; Value Stream Mapping [13, 14], Kaizen [15], Kanban [10]; can be implemented without any major investment in automation. Kaizen is a Japanese term that means continuous improvement. Kaizen Blitz or Kaizen Events are a common activity on the journey towards leanness.
where a group of trained lean experts target and review one or more production processes in the quest to identify opportunities for improvement. Value Stream Mapping is a technique used in Lean Manufacturing that maps the flow of material and data, and associated time requirements. It begins from the initial supplier to end at the customer for a given business process. It can be used to define improvement areas and sources of waste. Kanban is a method of flagging up the need for more parts on an assembly line by using signal cards at workstations, first used at Toyota and associated with the Just-In-Time technique. However, the implementation and deployment of these tools is not in itself Lean.

Womack and Jones [4] identified the five basic principles underpinning Lean Thinking as:

1. specify value by product;
2. identify the value stream for each product;
3. make value flow without interruptions;
4. pull value from the manufacturer; and
5. pursue perfection.

Bhasin and Burcher [5] identified that the Lean paradigm has two important requirements; technical requirements and cultural requirements. With respect to the technical requirements, they suggest that instead of focusing on one or two tools of lean in isolation that it is important that companies practice most, if not all, of the following: Kaizen (continuous improvement), Cellular Manufacturing, Kanban, Single-Piece Flow, Process Mapping, Single Minute Exchange of Dies (SMED), Kaikaku (radical activity improvements), Supplier Development, Supplier Base Reduction, SS, Total Productivity Maintenance, and focus on Value and the Seven Wastes (overproduction, waiting, transportation, inappropriate processing, inventory, unnecessary motions and defects). In addition to these technical requirements, Bhasin and Burcher [5] identified thirteen changes that are required in corporate culture to ensure successful implementation of Lean. These include, decision making at the lowest level, clarity of vision, strategy of change with clear communication of how the goals will be achieved, clear roles and responsibilities, develop supplier relationships based on trust and mutual respect, nurture a learning environment, focus on the customer, promote lean leadership at all levels with clear lean metrics, maintain the challenge of existing processes, maximise stability by reducing schedule changes, program restructures and procurement quality changes, access the fraction of employees operating under lean conditions, observe the proportion of departments pursuing lean and long-term commitment. Karlson and Ahlstrom [16] presented a conceptual model of Lean Production that argues that the paradigm ranges from Lean Product Development, to Lean Procurement, to Lean Manufacturing to Lean Distribution. Womack et al [3] used the term Lean Enterprise to describe the application of Lean Thinking outside the boundaries of the organisation.

Krishnamurthy and Yauch [17] summarise lean, therefore, as “a methodology of developing a value stream for all products that eliminates waste in waiting time, transport, inventories, and defects, and focuses on a level production schedule”.

3. SCHEDULING

Production scheduling is a critical activity in manufacturing. It concerns the distribution of scarce resources, usually machines, to tasks over time [18]. Scheduling is defined as a decision making query that entails optimisation of one or more scheduling conditions. Due to it combinatorial nature, scheduling problems are computationally very intricate and complicated to solve. Therefore, it is not always possible to find the best possible solution in a reasonable time frame. Assortments of heuristic methods have been developed in order to find near-optimal solutions in comparatively short periods of time. However, often heuristics applied in practice are dispatching rules that have minimal computational complexity and are simple to implement [19]. Production scheduling is important to manufacturing organisations for a number of reasons. J Younger a pioneering author in the field of scheduling had the following view on the matter:

“Well-organized and carefully executed work routing, scheduling and dispatching are necessary to bring production through in the required quantity, of the required quality, at the required time, and at the most reasonable cost.” [20]

Cost objectives, quality targets, delivery concerns, and quantity goals are the key elements and evident reasons why scheduling is performed by organisations. It also provides the basic background for the formulation of mathematical structures and computer systems architectures that simulate and generate schedules.
In most situations a production schedule will never be executed precisely. Disruptions are a certainty and modifications are inevitable to allow execution, and perhaps facilitate improvements in dealing with the situation encountered. This process of altering the original schedule to handle disruption from uncertain varying factors is generally referred to as rescheduling or reactive scheduling. Manufacturing operations can encounter a wide range of uncertainties. Therefore, the objectives of scheduling are to accommodate and anticipate these uncertainties before they occur or have a recipe to counteract them [21]. Different sources of uncertainty are inherent in real life production scenarios. The interruption that has been most often dealt with in the literature is machine breakdown. Scheduling research has so far been unable to properly address the general issue of uncertainty, making the impact of scheduling research less influential on industrial practice [21]. However, uncertainty is not the only dimension in scheduling, and research efforts need to be redirected to look at different aspects that effect scheduling.

3.1. Scheduling in a Lean Environment

Production planning in a lean environment requires smoothing out the peaks and valleys in the production schedule. This aids to maximise the utilisation of the production facility. Levelled production aims to run a constant quantity of all the operations, hence without the levelling system, there would be great difficulty in dealing with uncertain demand fluctuations. Unless there is a surplus of labour, capacity and large quantities of inventory, the concept of a levelled schedule is required. The benefits of this constant production and levelled schedule are reduced overall waste. This can be in the form of less operators standing ideal while waiting for work, or machines and tools that require high investment sitting unused [22]. In order to gain a better understanding on how Lean effects scheduling, it is worth while exploring the main tools and concepts that may have an impact [23].

3.1.1. Takt Time:

Takt time is the basic rate of production, also referred to as the drumbeat for the process of production. Takt time is equal to the operating time available divided by the customer’s required quantity, and uses the language of supply and demand, where it attempt to allow for the supply to meet or even exceed the demand in order to ensure that the customer order is fulfilled and avoid disappointment. Takt time is usually calculated prior to generating a schedule, the rest of the operations have to be aligned with the Takt time in order to avoid delays or shortages. However, instances where a production facility is faced with uncertainties such as the arrival of urgent orders, unpredictable machine breakdown or resource shortages may have an impact on the Takt time calculated. In such cases the Takt time needs to be recalculated incorporating remedial actions in order to revamp the schedule.

3.1.2. The Pacemaker Process and Heijunka:

When scheduling in a lean environment, the pacemaker must be initially identified, as this operation will determine the pace of the rest of the value chain. The pacemaker is usually a critical operation with limited recourses or capacity. Scheduling at this one point will result in a pulling effect on work from upstream processes and flowing product through the subsequent processes to the customer. The scheduled volume and product mix at the pacemaker typically corresponds to what is known as the master schedule. Master schedules are established in synchronisation with the Takt time for all items that go through the pacemaker process. The entire system depends on an elementary lean principle of levelled production called Heijunka [2, 22, 24, 25], which involves the levelling of production by both volume and product mix. Products are not manufacturer in accordance with the sequence of customer orders; rather Heijunka calculates the total volume of orders in a period and levels them out so the same amount and mix are manufactured each shift/day. Small-scale Lean organisations use spreadsheets to schedule their production in order to create Heijunka. However, IT systems are a crucial addition for most organisations and yield significant benefits. With the addition of the Internet, this has exploded its potential [26]. The spreadsheet approach can be helpful in a pilot context or small scale entities, but it is questionable whether spreadsheets are a scalable technology in larger organisations, as it’s known that they encounter issues with data reliability when used in isolation. Hence, there is a requirement to enable the scheduling method to integrate with an organisation’s Enterprise Resources Planning (ERP) or Supply Chain Management (SCM) systems.

3.1.3. Finished Goods Supermarkets:

Producing to customer orders only is, when possible, the best practice. However, producing a constant small inventory maybe more desirable than flexing plant resources to meet day-to-day order variations. In a lean operating plant, products are often produced to a buffer called the finished goods supermarket, rather than directly to customer
orders. The stability and reliability of customers’ ordering frequency will mandate the size of the supermarket required. Another dependent factor for the size of the supermarket is the replenishment time for the finished goods.

3.1.4. Pull to the Customer / Kanban:

A common standard in lean manufacturing is the principle of pull, which emphasises only replenishing what is used. Pull replenishment works on the basis of segmenting the Work-In-Process (WIP) or finished goods supermarkets into equal units referred to as Kanbans. Once a certain predetermined quantity of Kanban units is consumed, a signal is generated to indicate the requirement to schedule production or authorise a new order for replenishment. In order to allow for a Kanban system to be productively implemented, a few prerequisites are needed:

1. Demand for the items must be steady and continuous.
2. Replenishment time must be relatively short in comparison to order lead time.
3. Raw material needs to be readily available to allow immediate production to commence once a signal is generated.

When demand is not steady and continuous, quantitative modelling techniques such as discrete event simulation or queuing network models can provide decision makers with the means to optimise the parameters of the system, such as the number of items in Kanban bins or the number of Kanbans to be allocated to workstations.

4. Case Studies

To illustrate the benefit, if not the need, in introducing customised Lean tools to assist scheduling in complex production environments, two case studies are reviewed in this section. The case studies provide insights on the benefits of using analytical models to support scheduling in an integrated push-pull environment and in a cellular manufacturing environment, respectively.

4.1. Case Study 1: Scheduling in an Integrated Push-Pull Environment

The first case study involves a global supplier of eye healthcare products. In particular, the division of focus here is dedicated to the production of contact lenses. Contact lenses are produced over a two-step process in standard sized batches. The initial step involved the production of moulds for the lens. For each contact lens two moulds are required. The second step entails the injection of lens material into the mould cavity to produce the contact lens. The associated complications with this procedure are well identified. The moulds have a specific curing time to stabilise their structure. In addition moulds have a relatively short shelf life wherein they must be consumed to ensure the quality of the lenses. The moulds are destroyed during the second step and cannot be reused.

The company was aiming to overhaul their contact lens operation to accomplish higher value by reducing waste. This is a typical Lean Kaizen innovation. An ideal way to achieve a reduction in waste is by striving for zero inventories. However this was not considered possible in this case, due to the curing times involved. Hence, a reduction in the buffer stock between the two processes was sought. During that process, the company still emphasised the importance of maintaining the robustness of the system; problems with the production of moulds should not have immediate impact on the production of lenses. Engineers at the facility were unable to predict the consequences of changes their buffer strategy, and hence could not adequately evaluate the trade-off between balancing inventory quantities and maximising service level. They required a quantitative model to represent system and simulate different scenarios.

For the purposes of developing the model, the main area of interest was bound by the two production processes and the intermediary buffer mentioned above. In order to apply Lean scheduling techniques, Takt time is normally the initial computation, as the rest of the operations have to be aligned with this rate in order to avoid delays or shortages. However, in this instance, the production facility is faced with a variety of uncertainties such as the arrival of urgent orders, machine break downs and possible resource shortages, resulting in difficulties to determine a Takt time. Therefore, the factory here does not explicitly operate on the Takt time basis. Weekly forecasts are used to generate an efficient weekly schedule to produce the lenses. The schedule accounts for machine maintenance and also quarterly forecasts. The latter enables some smoothing to take place, distributing the load of heavier periods over less loaded periods. The model, therefore, utilises an elementary Lean principle, Heijunka (levelling production).
The intermediary buffer operates on a pull principle; this is a result of the second production process being the pacemaker. The second production process is the pacemaker due to resource and capacity limitations at that operation. To assess the buffer’s capacity, different Kanban levels were used to determine adequate batch sizes of moulds required to produce a certain number of lenses. Depending on the replenishment level being tested, it can either operate in a tightly coupled fashion (a minimal buffer is used, and stocks are replenished as soon as moulds are taken from the stock) or more loosely coupled (the buffer has enough moulds for several batches, and only after a number of batches are taken out, the moulds are replenished). This integration of both Push (in this case of lenses to the customer) and Pull (of moulds to the casting stage) is proposed in the Lean literature by e.g. Hodgson and Wang [27, 28] and Cochran and Kim [29]. Both place specific significance on requiring quantitative analysis to determine the parameters of such integration. Here, a form of quantitative modelling called Constraint-based Scheduling was used to perform the analysis [30].

The model had sufficient levels of functionality in statistical analysis, which were used for determining the most suitable levels of stock to be held at the buffer and also display the effect of different levels on the resulting schedule. This helped the engineers in weighing the savings gained by smoothing and quicker casting against higher stock levels for different scenarios. Due to the flexibility of the tool, additional scenarios were investigated. These included the addition and removal of machines to and from the facility, and also evaluating alternative materials to use for producing the moulds. In essence, this tool has provided the company with enhanced visibility of the scheduling process and improved knowledge of the scheduling consequences on modifying factory operations, concluding to be an invaluable tool in the company’s Lean journey.

4.2. Case Study 2: Scheduling in a Cellular Layout

The second case study deals with Alcatel-Lucent, a global telecommunication giant. In particular, the focus is around their base station fabrication facility. As these wireless infrastructures vary widely in configuration, the company practises a build-to-order strategy. The fabrication of these base stations consists of three major processes; assembly, wiring and testing. Due to the complex nature of the wiring and testing procedures, the company decided to implement a Lean driven cell-based workstation layout for Wiring and Testing. Dedicated cells were designed to handle limited number of product groups. The intention of the restructured shop floor layout was to improve the wiring quality and reduce the testing times. However, as a result of the dedicated work cells, the scheduling pattern will have a direct impact on the workload levels of each station. Planning errors can result in starving certain cells, or tail-backing already overloaded cells. Hence, the use of an analytical scheduling tool was required [31].

The site was already implementing a number of Lean tools and practices before reshaping into a cell-based layout. Working in a synchronised fashion tailored the production to a predefined Takt time, allowing the company to produce in a Just-In-Time mode. The principle extended over the entire production line, with each step ideally handing over goods as the succeeding station frees up. Workload durations are standardised across the assembly lines, but due to the reconfiguration, items will spend multiple durations of the beat at each of the new cells. Therefore, one assembly line can be feeding multiple cells. Natural production cycles at the plant have shown patterns of low demand early in quarter and peaks nearing the end of each quarter. This prompted implementing a levelling regime to smooth the production, however resulting in a quantity of orders being produced ahead of schedule.

The quantitative scheduling support tool developed here was able to account for most of the variables involved and for resource limitations. The final results from the model where localised smoothing at each cell and a reasonable levelled production rate of the assembly stations feeding in. It is difficult to display the direct benefits of the scheduling tool, as it converged into the overall new style of production at the facility. However the engineers have observed significant improvements in the manufacturing interval, work in process inventory, first test yields, head count, and quality of the delivered performance. Although these benefits are mainly achieved because of the reformed cell layout, the scheduling tool is crucial in realising its full potential. As a result, the company has considered extending the cell-based shop floor concept to other areas of the factory, and obviously continue to avail of quantitative modelling as part of the extension.

5. Discussion and Conclusion

The Lean manufacturing toolbox has been standardised to its pure elements, and will at all times require adjustment and customisation to the individual organisation. The tools themselves are not that complicated to use, the difficulty resides at choosing the right tools for the situation being dealt with. Implementing Lean should not be
managed as a project, but rather a new way of conducting business. The first improvements will usually be soon visible, but companies should not have high expectations for the first months. Lean takes time and demands resources to become a success. This is not referring to high capital investment, but when attempting to change the whole organisation’s way of working, this will take time.

Scheduling and planning are important for an enterprise to foresee and be able to control or adapt for its future. Anticipating what may take place, allows management to modify or adjust their actions in order to control the outcomes. There are many commercially available tools for use in the scheduling effort. Some are quite generic, but others concentrate on a specific business model. However, most of them share some fundamental similarities that include cost and resource utilisation as primary targets. This involves the delivery of orders to customers in the shortest feasible time span and at a price margin that is accepted by both manufacturer and consumer. Scheduling operations in a Lean environment, therefore, plays two important roles. Firstly, it has a role in developing improved strategies for dealing with chaotic uncertain demand. The aim is to reduce variability by using Kanban supermarkets, finish-to-order or make-to-order strategies in order to dampen the impact of fluctuating orders received at the plant. The second role of scheduling in a Lean environment is the levelling of the schedule, in order to allow the plant to balance production to a constant beat (Takt time) and not over-stress or underutilise the facility and resources available.

Scheduling can greatly benefit from the addition of lean principles; this was demonstrated in both case studies presented in Section 4. The use of quantitative modelling added the advantage of allowing the user to forecast the outcomes of such implementations and helped in supporting the decision making process. Thus, it is a significant addition to the lean toolset as it allows predicting the outcomes of strategic decisions and aids the rescheduling process in the event of any interruptions. The application of quantitative modelling techniques to support decision making in lean environments, in particular with respect to scheduling, provides the possibility to employ statistical analysis to a much greater advantage than demonstrated in the above case studies. It allows the decision makers to account for uncertainties, more significantly, multiple uncertainties simultaneously. This is done by allowing decision makers and users to tease out relationships and variables that might have been masked by other factors. The systematic nature of statistics can ease efforts at replication and extension; however, there are also disadvantages. It could encourage engineers to act as if they can ignore whatever they have yet to learn how to measure, which can lead to wrong conclusions and impair the research target.

The important point to note is that Lean should be seen as a direction, rather than as a status to be achieved after a certain time period. Therefore, the focus should lie on what the organisation seeks to achieve, and not what each Lean tool can achieve for them. It should also be noted that not all the tools, in the Lean toolbox, are adaptable to any situation. There could be instances where conflicting signals are sent. Thus, with this in mind, the application of such tools as quantitative modelling should be in a supportive role for practitioners to ensure that they are not diverging from their initial goals.

ACKNOWLEDGEMENTS

This research is part funded by the Embark Initiative from the Irish Research Council for Science, Engineering and Technology (IRCSET), the National Development Plan (Grant Reference Number: RS/INTEL2/07), and the Science Foundation Ireland (SFI; Grant Reference Number 08/RFP/CMS1711).

The case studies have received support from SFI (Grant Reference Numbers 00/PL1/CO75, 03/CE3/I405 (Centre for Telecommunications Value-Chain-Driven Research) and 05/IN/1886), and Enterprise Ireland (Grant Reference Number IP/2003/160).

REFERENCES


A Fundamental Mindset that Drives Improvements towards Lean Production

Yuji Yamamoto¹* and Monica Bellgran²†

¹ School of Innovation, Design, and Engineering
Mälardalen University
Eskilstuna, Sweden

² Haldex AB, Sweden, and Adj. Prof. at School of Innovation, Design, and Engineering, Mälardalen University, Eskilstuna, Sweden

ABSTRACT

Even with the recent development of how to apply lean production, many manufacturers still fail to succeed in lean transformation especially in the area of cultural change which is the key source of advantage of lean as well as the key of the sustainability of the transformation. Academic and practitioners increasingly recognize that contingent nature is required in order to achieve lean transformation with cultural change. This gives a view of lean transformation as a complex process where numerous improvement activities are conducted in an arbitrary manner depending on the context of company and individuals. Even so, is there any central thinking that drives various improvement activities during the transformation? One and half years of the author’s participant observations of two ongoing lean transformations at two Swedish manufacturing companies facilitated by a Japanese consultant who has 25 years of experience of Toyota Production System identifies a fundamental mindset that can be a central driver of the whole lean journey. The purpose of this paper is to present this fundamental mindset. The identified mindset is based on the belief—"every improvement starts from the needs" and focuses on how to create a need for improvement towards lean. An operative way of practicing the mindset and a way of institutionalizing the mindset into an organization are also suggested. Later, the potential and the risk of practicing the mindset are discussed and finally conclusion is drawn.

1. INTRODUCTION

Lean production is one of the most influential new paradigms in manufacturing. Its shop floor tools, techniques, and philosophies are widely applied in manufacturing industry. Lean production is originated at Toyota Motor Corporation and other Japanese manufactures. However, the world’s attention to lean production was limited before the publication of the book *The Machine That Changes the World* [1], in which also the term ‘lean production’ was introduced. In the early period of lean awareness, many of the manufacturers’ efforts was focused on the emulation of shop floor techniques of lean, such as 5S, flow production, small batch production, single minute exchange of dies (SMED), standardized work, and Kanban, and they found it difficult to introduce and sustain them. Later, the necessity of organizational cultural and mindset change in lean application is denoted by several authors [2] [3] [4]. Womack and Jones [3] identify the importance of “thinking” in lean production and summarize the five principles of “lean thinking” as a dependable guide for lean transformation. These principles involve the identification of customer value, the management of the value stream, developing the capability of flow production, the use of “pull” mechanisms to support flow of materials, and the pursuit of perfection through reducing to zero all forms of waste in the production system. Linker [4] further maintains that lean thinking based on the Toyota Way involves a far deeper and more pervasive cultural transformation than the application of set of lean tools, and he presents fourteen management principles as the foundation of Toyota Production System (TPS). The mentioned authors and many other advocates of lean production commonly agree that the goal of lean transformation is to achieve competitive and adaptive manufacturing with the culture of continuous improvement and organizational learning.
Although the recent development of how to apply lean production, many manufacturers still fail to succeed in lean transformation especially in the area of cultural change, even though it is the key source of advantage of lean as well as the key of the sustainability of the transformation. Previous experience in industry shows that it is difficult to achieve cultural change only by introducing shop floor lean techniques. Following prescriptive lean transformation steps, for example “lean leap” by W. Omack and Jones [4], do not always guarantee to bring about the cultural change. Lean transformation accompanied with cultural change seems to go through more complex process. The prescriptive approach or plan-based approach to lean transformation is criticized by Hines et al. [5], claiming that the contingent nature is required to apply lean thinking which is a general misunderstanding of lean application. Drew et al. [6] also maintain that lean transformation is not a project but a journey, stating “it may be tempting to turn these phases into a project plan or a process to be followed, the reality of the journey is not like that. There is no ‘right’ way to approach a lean transformation”. This perspective gives a view of a lean transformation consisting of numerous improvement activities where different lean techniques and principles are applied in an arbitrary manner largely depending on the context of company and individuals. Even so, is there any dependable guide that acts a central driver of the whole journey? Such driver provides a consistent way of driving improvement activities towards lean production, provides a coherent meaning of how to use different techniques and principles, and also gives a consistent way to stimulate production staffs and operators for collective learning.

One and half years of participant observations of two ongoing lean transformations at two Swedish manufacturing companies facilitated by a Japanese consultant who has 25 years of experience within TPS indicates a fundamental mindset that can be a central driver for the whole lean journey. The purpose of this paper is to present this fundamental mindset. How to practice the mindset as well as how to institutionalize the mindset in to an organization are also suggested. In the next chapter, the research method is presented. Then, several improvement episodes observed at the two manufacturing companies are introduced. These episodes are basis of the theory formulation in this paper. After the fundamental mindset is presented, discussion and conclusion are made.

2. RESEARCH METHOD

The fundamental mindset presented later in this paper is extracted from case studies of two ongoing lean transformations at two manufacturing companies in Sweden. One of the two companies (company A) is a medium sized company with about 150 employees. They produce precision casting goods for several industries such as automotive, industrial equipments, and infrastructure. Product variation is about 600. Before they started the lean journey in September 2007, they had done little improvement activities the past 20 years due to their niche market position with less competing pressure from other competitors. With their poor operational performance, which was the main reason of their poor profitability, the president was eager to initiate a drastic change. The other company (company B) is also a medium sized manufacturing company with about 130 employees. They produce electrical products mainly for infrastructural industry. This company is originally a R&D driven and little attention was drawn to the production. Due to the increasing production volume together with their great concern about their production quality, the managements wanted to apply lean production to their shop floor operation.

The lean transformations in those two companies have been facilitated by a Japanese consultant since September 2007. He had worked for one of the Toyota’s supplier companies in Japan focusing on TPS for more than 25 years. He retired from the company in 2005 and now continues to instruct TPS at a large number of companies on a global basis. In total, he has instructed TPS at more than 150 companies and 2800 persons globally. His consultation style is the one commonly practiced at Toyota and its related companies: going to shop floor, seeing and analyzing operation carefully, and suggesting improvement. A lean coordinator or production manager is required to accompany him during his consultation at the specific company. The focus is improvement of operation but the range of the change is often extended to whole company since change of management structure and their mindset is of the critical importance in lean transformation.

The research study is carried out by the main author’s participant observation. The author currently participates in these two lean transformations as a translator as well as an assistant of the Japanese consultant. The period of participation so far has been one and half years at company A and a half year at company B. Evidence is collected in various ways: through author’s experiences gained from actual improvement activities, discussion with the president, production managers, group leaders, engineers, and operators, and frequent consultation with the consultant that especially helps to understand the thinking behind of his behaviors, decisions, and actions during his consultation. Discussions, comments, events and thoughts considered important or relevant documented. Participant observation has a distinct benefit of the ability to perceive reality from the viewpoint of someone “inside” rather external to it, but it has also the risk of potential bias produced [7]. The author’s position as
a translator and an assistant consultant makes it easier to observe the events from third perspective and to limit direct involvement of improvement activities.

3. IMPROVEMENT EPISODES DURING THE LEAN TRANSFORMATIONS

During the participant observation at company A and B, the author observed a number of improvement activities. Many of these improvement episodes are used as data to formulate the fundamental mindset described in chapter four. In this chapter, we show some of the representative ones. The episodes are observed at the different points of time during the transformations. Most of the episodes are related to the consultant’s suggestions and advice, some of which were implemented immediately by the companies while others took several months to be understood and implemented. Some of them still have not been implemented.

3.1 EPISODES AT COMPANY A

The general direction of the improvement at company A was to reduce work-in-process (WIP), since the company was in the negative spiral of poor operation caused by increasing WIP. At the beginning of the journey, they had a chronic problem of delivery delay. To offset the risk of delay, production control section started to produce goods as much and early as possible. This increased WIP and lead time. It also increased process complexity that made actual WIP and lead time even longer.

(Episode a) At one production process, the Japanese consultant thought that there were too many buffer stocks. It was because they produced weekly batches. After a quick investigation showing that it was possible to produce with daily batches, he suggested to remove the buffer stock completely except the amount needed for the daily batch. The shopfloor supervisor of the process and the operators showed confusion and unwillingness to the change because they thought it was impossible. The consultant however told them to change anyway, saying that they would surely figure out how to manage with this amount of buffer stock.

(Episode b) At another process, the consultant found some pallet shelves that had too much space for the buffer stock. He requested to remove half of the shelves immediately. He said “half of the space is enough for the future buffer level. It is better to remove already now because then people have to think of a way not to produce too much. People like to place things as soon as they find space.”

(Episode c) Production had delivery delay constantly for many years but production did not see it as a serious problem. Production was controlled by achieving weekly volume which was calculated from an annual volume based on their prognosis. Meeting the volume was more concerned than meeting the delivery date. The Japanese consultant instructed them to change the way of controlling production from volume control to delivery accuracy control. He required the production manager to initiate delivery delay follow-ups and to do it everyday. Similar follow-up system was implemented at each production processes. During the follow-ups, the consultant repeatedly stressed that this was not to “follow-up” the delays but to identify the causes that hindered to achieve 100% delivery accuracy.

(Episode d) During all improvement activities, the consultant repeatedly said that everyone must think from the customers’ perspective. He continued that one must not accept any excuses that are attributed to internal reasons. For example, reasons of delivery delay such as lack of personal, machine breakdown, and internal defects have nothing to do with the customers. He emphasized that one must think how to solve these issues instead of excusing oneself.

(Episode e) There was a tension between the two departments; production planning and production. Production planning felt that production did not respect following the plan, and production felt that many production orders based on prognosis and frequent priority changes of the orders caused chaos at the production. The consultant told production planners that as far as they tried to reduce the risk of delivery delay by starting more than necessary and earlier than necessary, the competence of production would never be improved. They were advised to try to start only confirmed orders and moreover to start as late as possible. With consideration of actual value added time of their products, the consultant speculated that lead time could be much shorter. He instructed them to reduce the lead time by 30% or 50% in their planning system immediately and to start production orders later in accordance with the shortened lead time.

(Episode f) Production has a process where they receive casted goods from the casting process and repair them and/or finish the surface of them. The process is called “finishing process”. The company employed many operators to this process because of the high casting defect rate. The casting process, however, had little awareness of this
quality issue. The consultant told the finishing process that if a batch from the casting process contained many defects the finishing process should reject a whole batch and return it to the casting process and require them to reproduce or repair it, just like their customers usually do to the company. If the casting process did not face the consequences of producing defects, they would never try to improve their process.

### 3.2 Episodess at Company B

At the beginning of the journey, company B had much higher delivery accuracy and its production floor was less chaotic than company A. In the context of improving quality, initial focus of the improvement was to create a flow and implement standardized work.

**(Episode g)** Already on the second day of the journey, an assembly section initiated layout change to create flow production. Then, the consultant instructed the production manager to carefully observe how operators assembled the products. He said “layout change is just a first step. Now, observe the process carefully and find any factors that disturb the repeatability of the operation. The more disturbances, the higher the risk of quality problem. The disturbance factor can be lack of assembly instructions, poor product design for assembly, insufficient operator training system, inappropriate fixture or jig, inadequate position of parts feeding, malfunctions of testers, lack of parts, defect parts, supervisors skills, and many other reasons. Lead time can be shortened by this layout change but identifying and correcting all those disturbances is the main reason of this layout change.”

**(Episode h)** After changing to flow production, the consultant requested to the assembly section to match the rate of customer demand with the rate of production based on the principle of Just-In-Time, producing just what is needed, just when it is needed, and in just the amount needed. Matching the rate of customer demand and rate of production is called producing with “takt time”. The production manager, however, wanted to assemble the products as early as possible in order to save some safety days before the actual delivery date. The consultant explained to the production manager that it is difficult to see the minimum resource needed for the assembly without producing with takt time. He added that as far as there is extra time and resource available the assembly section would not feel the need of improving all the disturbances as mentioned in Episode g.

**(Episode i)** After changing to flow production at the assembly section, introduction of Andon was discussed. Andon is an alarm signaled by operators when they find an abnormality. When signalled, the leader of the area has to go to the place of the occurrence immediately and solve the abnormality right away. The production manager was willing to introduce it because he thought many quality concerns might be brought up to the surface by introducing it. The consultant agreed to the possible benefit of using Andon but doubted sustainability of its application. He said that the assembly section still did not produce with takt time and therefore the operators would not feel the need of using an urgent call even when they faced the disturbances.

These episodes above may appear to have little in common, however, a consistent thinking about improvement can be identified from the consultant’s different comments and behaviours.

### 4. A Fundamental Mindset that Drives Improvements Towards Lean Production

A fundamental mindset can be derived from the improvement episodes observed during the case studies. In this chapter, this fundamental mindset is presented. Later, an operative way of practicing the mindset is suggested, and then we suggest how to institutionalize the mindset into an organization.

#### 4.1. The Fundamental Mindset

Through the analysis of the various comments and behaviors made by the consultant, it is possible to identify a fundamental mindset that gives a coherent reasoning to all of the comments and behaviors made by the consultant. This mindset acts as the central force that drives improvement activities at shop floor towards lean production. This mindset is summarized as “occasionally by force, create a situation where people have no choice (or little choice) but to feel the need of improvement. The situation is such that it brings different wastes and problems up to surface. Through letting people solving these wastes and problems one-by-one, the performance of the operation as well as the capability of individual/organizational learning are improved”.

This mindset is closely related to one of the beliefs of Taichi Ohno, who is one of the founders of TPS. From his plentiful experience of creating and developing TPS, he concluded that every improvements starts from the needs [8]. He maintains that improvements without feeling the need of them tend to end with short life or to fail to yield...
A Fundamental Mindset that Drives Improvements towards Lean Production

benefits proportional to the investments made for the improvements. He also claims that provoking people to feel the need for improvements is the key for the improvements. It can be said that the fundamental mindset is a practical way of conducting Ohno’s belief.

An image of this mindset can be explained by using the “Japanese sea model” as show in Figure 1. This model is usually used to explain the concept of reducing stock as an improvement activity towards lean production. When the water level is high, the objects are hidden under the water. By reducing the level of it, the objects are brought up to the surface. Likewise, high stock level hides different problems. Problems such as lack of parts, producing defect parts, and machine break-downs are absorbed by the stock and do not affect the operation directly. Consequently, these problems are not likely to be recognized with the sense of urgency. When the stock level is reduced, however, the problems start to directly affect the operation. Therefore, they have to be solved immediately. In the fundamental mindset, the water level does not only represent stock level but also other parameters. The parameters are such that by changing them problems are brought up to surface and then people feel the need of solving the problems with the sense of urgency. The next section shows an operative way of practicing this fundamental mindset.

![Figure 1: The “Japanese sea” model](image)

4.2. AN OPERATIVE WAY OF PRACTICING THE MINDSET

The study finds that an improvement cycle with the fundamental mindset consists of four steps: Reduce, See, Think, Act. The first step is to trigger the need for improvement, and the other three steps are general problem-solving processes in continuous improvement, called Kaizen in Japanese. Each step is described below.

**Reduce:** This step is to change or to set a parameter by which people have little choice but to feel the need for improvement. Possible parameters are: level of stock, space for stock (see Episode a, b), connectedness of flow (see Episode g), takt time production with minimum resource (see Episode h), standard in-process stock, lead time in production planning system (see Episode e), commitment to 100% delivery accuracy (see Episode c), no reception of defect parts from the previous process (see Episode f), or number of operators and staffs. Choice of parameter is largely context-based, but there is a general rule. The consultant recommends first to change parameters related to delivery then to proceed to change those related to quality and cost, because problems related to delivery tend to be treated as higher priority. The extent of change should be the one that provokes people to feel the need for change. Too little extent does not make people feel the need with the sense of urgency, while too much extent makes people discouraged. “Edge of chaos” – the edge between order and chaos where the creativity, growth, and use of self-organization are at the optimal [9] can be a good indication for parameter setting. The initiator of such parameter change has to have an authority to change, is able to take risk for the operation, and is able to see the operation from a holistic perspective. Therefore, the initiator is best suited to be the production manager. No lean co ordinators or production engineers is suitable to do this task. Much resistance can be expected from staffs, leaders, and operators, when initiating the change. The production manager needs to be persistent about the change, otherwise the whole point of this kind of improvement approach is lost.

**See:** This step is to carefully observe shop floor to identify the problems that are brought up to surface. Then the causes of these problems are analyzed. Eight guidelines for problem-solving in Kaizen suggested by Yamamoto [10] can be used (see Figure 2). “Repeat why when one sees the abnormalities” (see G3 in Figure 2) and “Do not blame operators but blame system or standard” (see G4 in Figure 3) are useful guidelines to understand the problem structure in this step. The latter guideline says that problems shall not be attributed to operators, but a root cause can be attributed to the leader ship of supervisor or managers. An image of how the emerged problems are linked to
different causes is shown in Figure 3. The person who shall operate this step as well as the rest of the steps depends on the maturity of Kaizen of the organization. The more mature organization, people at the lower level in the organization can operate the Kaizen process. In the case of company A and B, their maturities of Kaizen are low, thus the production managers have to drive the problem solving steps until the shopfloor leaders and operators learn how to perform the problem solving.

Figure 2: The eight guidelines for problem-solving in Kaizen, modified from Yamamoto [10]
Think and Act: The “Think” step is about coming up with the solution to the problems and the “Act” step is to implement the identified solution. In the think step, the guidelines of “use wisdom thoroughly before use money” (see G5 in Figure 2) and “Create temporary solution even if the optimal solution is unknown or takes time to be implemented” (see G6 in Figure 2) can be referred. In the act step, the guidelines of “initiate change immediately when a solution is available” (see G7 in Figure 2) and “initiate change even if there is an uncertainty, more improvements will be found after the change” (see G8 in Figure 2) can be referred. If a parameter change initiated by the Reduce step causes the status of edge of chaos at the affected area, people at shop floor aspire for an immediate solution. Then, lean techniques and principles or other support methods become useful help for solving the problems. In this way, people can learn the meaning of the techniques and principles better than when they are merely provided without people feeling the need of them. The more appreciated the solutions are, the easier they are kept in use.

4.3. INSTITUTIONALIZE THE MINDSET INTO AN ORGANIZATION

Many companies are use to secure their operations by increasing safety margins such as buffer stocks. They are also use to do improvements without risking the operations. For such companies, conducting improvements based on the fundamental mindset may appear to be a completely opposite way of improving the production. Both potential and risk of the new way of thinking must be first understood by the production manager and higher up in the organization. The potential and the risk of practicing the mindset will be discussed in the next chapter. Accepting the new mindset is difficult. The Japanese consultant’s experience and the author’s experience from the observations of the company A and B are consistent with the expression of Drew et al. [7]: “written documents and conversational presentations are not that helpful” and “it could take months or even years for the leaders of one company to decide to embark on the journey, whereas other might do so in weeks”. The production manager at one of the discussed company took nearly one year to understand the fundamental mindset, while the production manager at the other company took a month to do so. The Japanese consultant and Drew et al. [7] also agree that people tend to accept the new way of working only through experimenting it by themselves.
5. DISCUSSION AND CONCLUSION

While the contingency nature is required for lean transformation, the fundamental mindset presented in this paper provides a consistent way of driving improvements towards lean production. Moreover, from the perspective of the fundamental mindset, the purpose of using lean techniques and principles is clear and consistent. With the mindset, most of those techniques and principles, for example, flow, pull, Kanban, standardized work (takt time, working sequence, standard in-process stock), Andon, jidoka, visualization, customer focus, and perfection are used in order to bring the problems up to surface and to make them perceived as urgent.

As mentioned in the chapter four, the potential and the risk of practicing the fundamental mindset are necessary to be analyzed. Table 1 shows the comparison of two improvement approaches towards lean production: one is improvements with the fundamental mindset and the other is plan-based improvements, where a detailed implementation plan is made prior to its installation into the production. For the plan-based approach, the transformation process can be more systematic than the other, but the process is more rigid and difficult to change itself when contingent external or internal changes around production occur during the transformation. The largest weakness of the fundamental mindset approach is that it is more difficult to know exactly what outcome will be obtained after the improvements and when the desired outcome will be obtained. It is so especially at the beginning of the transformation. Perhaps this is one of the reasons that managers sometimes have difficulty in accepting to this less systematic approach. On the other hand, the largest benefit of this approach is that it includes the process of collective learning through actions and therefore it has a higher chance to achieve the cultural change of a learning organization. In this approach, improvements are often triggered by the managers but actual problem-solving is expected to be done by the leaders and operators on the shop floor. They face the problems with the sense of urgency. By trying to solve the problems and, if needed, being supported by the managers and consultants who provide adequate tools and methods to them, the leaders and operators can learn how to train their knowledge and skills. In the other approach, however, solutions are often already designed during the planning phase and given to the shop floor to be implemented. There, leaders and operators have less chance to learn how to improve their actions through better knowledge and understanding.

Table 1: Comparison of improvement approaches towards lean production

<table>
<thead>
<tr>
<th>Comparison items</th>
<th>Improvements based on the fundamental mindset</th>
<th>Plan-based improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformation process</td>
<td>Contingent but flexible</td>
<td>Systematic but rigid</td>
</tr>
<tr>
<td>Outcomes: what and when</td>
<td>Less certain</td>
<td>Certain</td>
</tr>
<tr>
<td>Behavior of improvement actors</td>
<td>Autonomous Managed</td>
<td>Managed</td>
</tr>
<tr>
<td>Much time spent in</td>
<td>Actions</td>
<td>Planning and negotiations</td>
</tr>
<tr>
<td>Involvement of people at shop floor</td>
<td>Learned to do</td>
<td>Told to do</td>
</tr>
<tr>
<td>Solutions</td>
<td>Created on the spot of problem occurrence</td>
<td>Already prepared in the plan</td>
</tr>
<tr>
<td>Chance of cultural change to learning organization</td>
<td>Higher because improvement process itself contains learning process</td>
<td>Less likely to be achieved</td>
</tr>
</tbody>
</table>
The analysis of the observation and theory formulation in this paper is made in the context of lean transformation at medium sized companies. At larger sized company, a more systematic aspect is probably needed. As Hines et al. [11] suggest, the application of policy deployment with consideration of various contingent factors can be a useful way to keep the balance of contingency and systematic aspect required for lean transformation. Finally, the paper implies that the fundamental mindset can be used not only for the collective learning but also for the stimulation of creativity which leads to an innovation in production. Innovation may need to be brought about if a production is already highly competitive using any possible best practices available in industry but is still required to reduce the water level drastically. The linkage of the fundamental thinking and innovation is an interesting area of further research.

REFERENCES

JIT Performance Measurement in a Specialist Packaging Manufacturing Environment

Y. G. Sandanayake* and C. F. Oduoza

School of Engineering and Built Environment, University of Wolverhampton, Shifnal Road, Priorslee, Telford, TF2 9NT, United Kingdom

ABSTRACT

Over the last decade, manufacturing environments have been challenged to radical implementation of typical philosophies such as just-in-time manufacturing (JIT), agile manufacturing, and rapid response manufacturing to achieve dramatic performance improvement. Literature review shows that there is no mechanism so far to identify the major JIT drivers and their impact on enterprise performance especially for small and medium (SME) non-automotive manufacturing environment. This paper presents a methodology for the systematic identification of the influence of key JIT drivers on performance in a food packaging manufacturing industry using a multidimensional performance measurement tool. The study has applied action research methodology and a novel approach to an SME that provides innovative solutions to the packaging industry. The outcome is a robust mechanism that enables the identification of key JIT drivers \((X_i)\) for their process with a resultant positive impact on operational performance \((Y)\). The generic performance measurement model developed here can also be applied to any other JIT enabled environment using the relevant organization specific JIT drivers to optimize system performance.

1. INTRODUCTION

Under intensive global competitive pressure, most companies around the world have applied innovative thinking to management and begun to examine technology that can lead to improved manufacturing flexibility, product quality and production cost [1]. Time-based competition is one of the most important recent trends in a business environment and Just-in-Time (JIT) philosophy plays an increasingly prominent role in the modern industrialized era [2]. JIT is an all-inclusive organizational philosophy designed to achieve high volume production using minimum inventory at the right time and based on planned elimination of all waste and continuous improvement [3]. Performance measurement system in a JIT environment must provide the requisite measures and control to support management decision making in terms of JIT strategies.

Empirical research so far involving the study of performance measurement in JIT production environments consists primarily of mass scale questionnaire surveys and case studies of specific organizations. More recently, some studies have also described simulations and mathematical modelling in JIT research [2, 4 and 5]. Mathematical modelling of JIT generally focuses on relationships between changes in various production factors and the corresponding specific production performance measures.

Quantitative information is helpful in implementing JIT production techniques and computer simulation can be a valuable tool in designing, implementing or changing JIT practices in a production system [4]. Computer simulation and modelling tools help to visualize, analyze and optimize complex production processes using computer animation to minimize the time and cost of a process [6].

To foster manufacturing strategies such as JIT, PMS may need to link JIT techniques with company goals, strategy, critical successes factors and key performance indicators (KPIs) [6]. Some researchers have assessed the impact of limited JIT techniques on aspects of productivity and performance in US firms without conclusive and substantiated results while other previous researchers have failed to quantify the effect of various JIT factors on performance [7]. The major shortcoming from previous research in this area is the lack of a comprehensive and

* Corresponding author: Tel.: (+44) 1902323944; Fax: (+44) 1902353843; E-mail: yasangee@yahoo.com
elaborate PMS to assess success, failure or impact of JIT practices on total enterprise performance. There is therefore, a void in the literature on JIT techniques for a suitable tool to measure its impact on performance, especially for small and medium non-automotive manufacturing environment.

This study will therefore introduce a robust, comprehensive performance measurement tool enabling a multidimensional assessment of the impact of JIT techniques on enterprise performance. The aim is to identify key JIT variables that drive performance using a performance measurement system such as the extended Balanced Score Card (BSC). The outcome from the study will provide a performance measurement model and implementation procedure for the successful implementation of JIT techniques in production environments.

2. PERFORMANCE MEASUREMENT MODEL AND IMPLEMENTATION PROCEDURE

Figure 1 illustrates the proposed conceptual model and inter-relationships between JIT drivers and performance. The left side of the model lists universal JIT techniques, which theoretically drive enterprise performance. Universal JIT techniques are divided into three groups: unique JIT practices, TQM related JIT practices and human/strategic oriented JIT practices. These three categories have similar fundamental objectives, which are to assist pull production, minimize waste and lead-time, and achieve continuous improvement. Unique JIT practices consist of pull system (kanban), line balancing, setup time elimination plans, JIT purchasing, level schedules, group technology, focused factory and inventory transportation systems. TQM related JIT practices include quality control activities, quality circles, value analysis, total productive and preventive maintenance, integrated product and process design and workplace organization plan. Effective communication, supplier integration, employee training, multifunction employee, innovation and investment plans, and other control techniques are variables categorized under human and strategic oriented JIT practices.

The right side of the model depicts performance measurement using the extended BSC as a , multidimensional and elaborate PMS to assess enterprise performance not only from the economic and financial standpoints but also based on the influence of the customer, supplier, employee, internal business processes, external environmental groups as well as innovation and growth perspectives [8]. In performance measurement model implementation, the integrated framework of JIT practices will provide relevant JIT techniques and the extended BSC will offer key performance indicators to establish cause and effect relationships between the most influential JIT drivers and measurable performance. The resultant performance measurement model will assist researchers and practitioners to take necessary actions to optimize JIT manufacturing performance in a continuous improvement exercise.

The conceptual model links key JIT drivers \( (X_i) \) and measurable performance \( (Y) \) through a mathematical model (i.e. \( Y = f(X_i) \)) with an aim to assist the user on the systematic identification of the influence of key JIT drivers on organizational competitive priorities.

Figure 1: Model Showing the Relationship between Operational JIT Drivers and Multidimensional Performance Measurement
The performance measurement model described here provides a generic, comprehensive and non-perspective template, which managers can apply to develop their own PMS matching the company strategy and priorities. This study applied design of experiments (DoE), simulation and mathematical modelling to establish the relationship quantitatively between JIT variables and operational performance in production environments. The performance measurement model implementation procedure is divided into a set of logical steps and a five-step implementation procedure is given in Figure 2.

Figure 2: Five-Step Performance Measurement Model Implementation Procedure

The application of the proposed performance measurement model and implementation procedure to a JIT enabled small and medium enterprise manufacturing environment is presented in the next section.

3. CASE STUDY: RISANE LTD.

The company selected for the application is Risane Ltd., which was founded in 2003 as a small and medium enterprise with the simple objective of providing innovative solutions to the packaging industry. The objective of the case study, which is based on action research, is to investigate the applicability of the proposed model to a JIT enabled non-automotive, small and medium production environment. Risane Ltd. is a fast growing company with 80 employees and it achieved a turnover of £3.7 million in the year 2006.

Risane Ltd. manufactures a range of packaging and other products including absorbents, bio-degradable, bio-compostable, microwaveable and protective materials for food packaging and for other applications such as boneguard products and medical products. The performance measurement model is applied to Packaging Type ‘A’ production process. This manually operated mixed model manufacturing process produces three types of products. The machines are located in three different building units and products are moved using pump trucks. Some products are only processed by one machine before being sent to a final customer, while the other products are processed by one machine and then sent to another machine or to an outsourced company for further processing before going on to the customer. The company does not have an in-house printing facility and hence printing is currently outsourced to an outside organization.

The Packaging Type ‘A’ production process layout (with photographs) and flowchart are shown in Figures 3(a) and (b) respectively.

Figure 3: (a) Packaging Type ‘A’ Manufacturing Process Layout and (b) Process Flowchart

The following section presents the application of the performance measurement model implementation procedure to the Packaging Type ‘A’ production process.
3.1 **Step 1: Determination of Key JIT Drivers of the Production Process**

Risane Ltd. has experienced long production process times for various reasons and Figure 4 depicts the detailed possible causes for this problem.

Ten major factors were identified using documents, observations and open-ended interviews with managers and operators. The critical factors affecting process time are positioning of machinery, machine idle time, setup time, machine integration, labour idle time, customer demand, line balancing, batch size, automation and outsource printing.

Relations diagram analysis was subsequently carried out to identify key drivers affecting process time (Figure 5). According to the relations diagram for Risane Ltd., “extended process time” was a major concern with ten causes (0, 10); “positioning of machinery”, “outsource printing”, “inadequate machine integration” and “high machine setup time” were identified as key causes with values (6, 2), (6, 1), (5, 2) and (4, 2) respectively.

---

**Figure 4: Cause and Effect Diagram Analysis for Process Time for Risane Ltd.**

**Figure 5: Relations Diagram Analysis for Identification of Key Factors Affecting Process Time for Risane Ltd.**
These four key JIT drivers affecting process time are now integrated into the customized performance measurement model (Figure 6) for further investigation.

The customized performance measurement model will now be tested and validated by conducting simulation experiments and AHP analysis to understand the effect of JIT techniques on performance of Risane Ltd.

3.2 **STEP 2: EXPERIMENTAL DESIGN AND MATHEMATICAL MODELLING**

The experiment is designed to identify and estimate the influence of the four key JIT drivers on process performance. In the simulation experiments, process time is calculated and evaluated. The four factor mathematical model developed for Risane is as follows:

\[ Y = a_0 + a_1A + a_2B + a_3C + a_4D + a_{AB} + a_{AC} + a_{AD} + a_{BC} + a_{BD} + a_{CD} + a_{ABC} + a_{ABD} + a_{ACD} + a_{BCD} + a_{ABCD} + \varepsilon \]  

Where, \( Y \): process time (KPI), \( a_0 \): intercept coefficient, \( a_1 \) to \( a_4 \): main effect coefficients, \( a_5 \) to \( a_{10} \): two way interaction coefficients, \( a_{11} \) to \( a_{14} \): three way interaction coefficients, \( a_{15} \): four way interaction coefficient, A, B, C and D: JIT drivers (as shown in Table 1) and \( \varepsilon \): error term.

### Table 1: Factors and Levels for the Experimental Design

<table>
<thead>
<tr>
<th>Factor</th>
<th>Key JIT Driver of Process Time</th>
<th>Variable</th>
<th>Lower Value (L)</th>
<th>Upper Value (U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Setup Time Elimination Plans</td>
<td>Setup Time</td>
<td>Ten Minutes</td>
<td>One Hour</td>
</tr>
<tr>
<td>B</td>
<td>Group Technology – 1 (Machine Integration)</td>
<td>Number of Work Stations</td>
<td>Three Work Stations</td>
<td>Four Work Stations</td>
</tr>
<tr>
<td>C</td>
<td>Innovation and Investment Plans</td>
<td>Printing Option</td>
<td>In-house Printing</td>
<td>Outsource Printing</td>
</tr>
<tr>
<td>D</td>
<td>Group Technology – 2 (Cellular Manufacturing)</td>
<td>Location of Work Station/Machinery</td>
<td>Single Location</td>
<td>Several Location</td>
</tr>
</tbody>
</table>

3.2.1 **SETUP TIME ELIMINATION PLANS (SETUP TIME)**

The lower setup time is defined as ten minutes when there is a change between two batches of a similar product family. The upper value is defined as one hour when there is a change between two batches of distinct families.
3.2.2 **GROUP TECHNOLOGY – 1 (NUMBER OF WORKSTATIONS/MACHINERY)**

The Packaging Type ‘A’ production process currently consists of three flexible process machines (laminator, sheeting machine and Sammy machine) and a packing bench. The upper number of workstations was therefore defined as four. During discussions with the Managing Director (Innovations), the researcher identified the possibility of merging the laminator and sheeting machines. The lower value was defined as three workstations, where there are two process machines (integrated machine and Sammy machine) and a packing bench.

3.2.3 **INNOVATION AND INVESTMENT PLANS (PRINTING OPTION)**

Risane Ltd. currently outsources printing to an external printing company. The Managing Director (Operations and Sales) expressed willingness to invest in new printing facilities during the interview sessions. Thus the upper value was defined as outsource printing, while the lower value was defined as in-house printing.

3.2.4 **GROUP TECHNOLOGY – 2 (LOCATION OF WORKSTATION)**

The machines are currently located in three different buildings but the managers suggested that all machines could be positioned in one location. Hence, the upper value was defined as several locations and the lower value as a single location.

3.2.5 **PROCESS TIME**

Process time is the total time taken to produce one assembly schedule. It is comprised of setup time, in-house activity time, move time within factory floor, loading/unloading time, transport time to/from printing press, printing time and inspection time.

3.3 **STEP 3: MODELLING AND SIMULATION EXPERIMENTS**

Data obtained from plant observations and managers’ tacit knowledge were used to develop simulation models using the ProModel simulation software. The assembly time at each station was determined using a stopwatch. The average of ten stopwatch readings was taken as an assembly time for each activity. The Auto::Fit function of the Stat::Fit software is used to calculate the appropriate continuous or discrete distributions to fit the input data. Simulation experiments were conducted for two batches of 3,135,000 products; 16 experiments were simulated, each with ten replications. Figure 7 presents a typical screenshot of experiment with ProModel.

![Figure 7: Typical Screenshot of Experiments with ProModel](image)
3.4 **Step 4: Simulation Result Analysis and Modelling**

Each experiment was simulated with ten replications and the maximum and minimum outputs for process time were selected for performance analysis. Table 2(a) shows the combination matrix for 2^4 full factorial design and replication output for process time, for two replications each in a Packaging Type ‘A’ production process. Statistical analysis and factorial fit for the experiments are presented in Table 2(b). Analysis of main effect, and two-way, three-way and four-way interactions were obtained at a 0.05 significance level (α).

Table 2: (a) Experimental Trials and Results for 2^4 Factorial Design and (b) Factorial Fit for Process Time

<table>
<thead>
<tr>
<th>Response</th>
<th>Setup time elimination plans (A)</th>
<th>Group technology – 1 (B)</th>
<th>Innovation and investment plans (C)</th>
<th>Group technology – 2 (D)</th>
<th>Process time (Hours) Replication 1</th>
<th>Process time (Hours) Replication 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>460.13</td>
<td>460.32</td>
</tr>
<tr>
<td>R2</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>460.33</td>
<td>464.95</td>
</tr>
<tr>
<td>R3</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>334.02</td>
<td>338.07</td>
</tr>
<tr>
<td>R4</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>333.61</td>
<td>337.13</td>
</tr>
<tr>
<td>R5</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>460.17</td>
<td>461.49</td>
</tr>
<tr>
<td>R6</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>459.05</td>
<td>461.63</td>
</tr>
<tr>
<td>R7</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>333.50</td>
<td>333.74</td>
</tr>
<tr>
<td>R8</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>332.70</td>
<td>333.45</td>
</tr>
<tr>
<td>R9</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>459.30</td>
<td>462.98</td>
</tr>
<tr>
<td>R10</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>459.79</td>
<td>461.65</td>
</tr>
<tr>
<td>R11</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>333.18</td>
<td>337.23</td>
</tr>
<tr>
<td>R12</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>332.76</td>
<td>336.29</td>
</tr>
<tr>
<td>R13</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>459.34</td>
<td>460.65</td>
</tr>
<tr>
<td>R14</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>458.24</td>
<td>460.80</td>
</tr>
<tr>
<td>R15</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>328.47</td>
<td>332.67</td>
</tr>
<tr>
<td>R16</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>331.86</td>
<td>332.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key JIT Drivers and Interactions</th>
<th>Coefficient</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Setup time elimination plans</td>
<td>Intercept</td>
<td>397.362</td>
</tr>
<tr>
<td>B: Group technology – 1</td>
<td>A</td>
<td>0.625</td>
</tr>
<tr>
<td>C: Innovation and investment plans</td>
<td>B</td>
<td>1.090</td>
</tr>
<tr>
<td>D: Group technology – 2</td>
<td>C</td>
<td>63.531</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>-0.070</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>-0.072</td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>BC</td>
<td>-0.365</td>
</tr>
<tr>
<td></td>
<td>BD</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>CD</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>ABC</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>ABD</td>
<td>-0.207</td>
</tr>
<tr>
<td></td>
<td>ACD</td>
<td>-0.206</td>
</tr>
<tr>
<td></td>
<td>BCD</td>
<td>-0.232</td>
</tr>
<tr>
<td></td>
<td>ABCD</td>
<td>0.071</td>
</tr>
</tbody>
</table>

According to Table 2(b), p values for main effects of group technology – 1 (B), and innovation and investment plans (C) on process time are both less than 0.05 (α). These two factors are therefore deemed statistically significant. All the other main and interaction effects are not statistically significant. For the case company, the effect of coefficient of innovation and investment plan show the highest statistical significance and has the greatest impact on process time. Moreover, process time decreased slightly when laminator and sheeting machines were grouped together. Although setup time (A) and group technology – 2 (D) are not statistically significant, process time reduces when setup time and location of machinery change from their upper to lower values. The aforementioned statistically significant factors are then built into the regression equation.

Based on the factorial fit analysis for the process time, equation (2) was derived, where, B: Group technology-1 (number of work stations) and C: Innovation and investment plans (printing option):

\[
\text{Process Time} = 397.362 + 1.09B + 63.531C
\]

(2)

The company can use the mathematical model described in equation (2) to predict process time by assigning upper and lower values to aforementioned factors.

3.5 **Step 5: Process Optimization**

The following conclusions and recommendations are aimed at optimizing both process efficiency and overall performance of the company:

- From the simulation studies, innovation and investment plans (printing facility) is identified as an extremely significant JIT driver and group technology – 1 (machine integration) is recognized as a fairly influential driver on operational performance. There are no interactions between key JIT drivers under the current manufacturing environment (Table 2(b)).
Operational performance can be improved (for example, process time can be reduced by approximately 27% on the Packaging Type ‘A’ production process) by installing an in-house printing machine.

Process time has been further reduced by integrating machines, providing the new integrated machine operates at a similar process rate compared to the maximum rate of the individual machines.

The company has now been able to improve the laminator (laminator + sheeting machine) while reducing labour cost, factory floor space (occupied by sheeting machine and material storage), unnecessary motion time, waste and defects by integrating machines.

Operational performance can be further optimized by arranging machineries in a work cell.

The existing machine setup times are high; however, the impact of setup time is hidden due to the high activity time at the laminator. Improving the laminator will reduce setup time between two different products by 85%.

4. CONCLUSIONS

The performance measurement model that linked key JIT drivers \( (X) \) and measurable performance \( (Y) \) through a mathematical model and a novel implementation procedure has been presented in this paper. It was aimed to assist researchers and managers in the systematic identification of the influence of key JIT drivers on organizational competitive priorities using a multidimensional tool such as the extended BSC.

The study applied the performance measurement model and implementation procedure to identify the impact of key JIT drivers (i.e. setup time elimination plans, group technology and innovation and investment plans) on KPI (i.e. process time) for Risane Ltd. This case study applied simulation and modelling using ProModel software and mathematical modelling to identify the impact of key JIT drivers on operational performance. It was found that the lack of innovation and investment plan practice had very high negative impact on process time. Improper implementation of group technology concepts such as machine integration and cellular manufacturing, and high machine setup times had a negative impact on operational performance (although not highly statistically significant). Consequently, it was recommended that the company should invest in new in-house printing facilities, integrate the laminator and sheeting machines, improve the new machine’s process rate and speed, arrange the machines in a cellular format and reduce the setup time of machineries in order to optimize the system performance.

The performance measurement model developed here can serve as a tool to enhance JIT manufacturing performance especially in identical small and medium non-automotive industry plants. The contribution to knowledge has been made by developing an innovative, easy to apply, robust methodology to enable industry practitioners optimize their processes and achieve higher productivity.

5. REFERENCES


JIT Performance Measurement in a Specialist Packaging Manufacturing Environment

Dynamic Workflow Planning in Holonic Manufacturing Systems

Fu-Shiung Hsieh

Department of Computer Science and Information Engineering
Chaoyang University of Technology
Taichung County, 41349 Taiwan

ABSTRACT

Planning of processes in holonic manufacturing systems (HMS) is driven by customers’ orders. Given an order with specific product demands, the holonic process planning problem is to find and organize a set of holons in HMS to compose production processes to fulfill the order requirements. A challenge is to design a mechanism to guide the holons such that the decisions made by the individual holons as a whole meet customers’ order requirements. Our objectives are to propose a methodology to develop a problem solver for the holonic process planning problem. To achieve these objectives, we combine multi-agent systems technology with Petri nets in this paper. We first propose architecture to solve the holonic process planning problem. We then propose a two layer contract net protocol to describe the negotiation between order holons, product holons and resource holons in HMS. To determine whether it is feasible for a set product holons and resource holons to meet the requirements of an order holon, we introduce Petri net models, which can be specified in Petri Net Markup Language (PNML) to capture the workflows and activities in product holons and resource holons, respectively. To enable a holon to discover the services provided by other holons, we propose a scheme for publication and discovery of holon services based on a FIPA compliant multi-agent system platform. Based on the service publication and discovery scheme, we develop a problem solver for holonic process planning and demonstrate the holonic process planning solver by an example.

1. INTRODUCTION

The manufacturing sector has been facing major challenges as it undergoes revolutionary changes fuelled by new demands from customers, global competition and technological advances [1]. Due to the dynamic characteristics of manufacturing environments, static control structures are not suitable anymore. The increasing versatility in product demands calls for a system architecture able to evolve in time [16]. Planning production processes with resource constraints, precedence constraints and timing constraints is a difficult task. Holonic manufacturing systems (HMS) [3] provide a reconfigurable, flexible and decentralized manufacturing environment to accommodate changes and meet customers’ requirements dynamically based on the notion of holon [19], an autonomous, co-operative and intelligent entity able to collaborate with other holons to process the tasks. In existing literature, there are many studies on HMS [4], [5], [7], [8], [10], [13], [14], [17]. Agent or holonic-based control architectures provides a distributed approach to overcome the limitations of classical planning and scheduling methods. There are many works on planning and optimization in HMS. Giebels, Kals and Zijm [4], [17] proposed the system architecture of a flexible manufacturing planning and control system, named EtoPlan, for concurrent manufacturing planning and control of activities and groups of resources in a manufacture-to-order environment. In [9], the authors presented an approach based on agent negotiation with an extended contracting protocol for supporting logistics and production planning, taking into account not only availability but also cost of logistic service providers. Leitão et al. studied on how should the production control structure evolve to adapt to changes have been made by [10]. Babiceanu, Chen & Sturges presented a solution for scheduling using the holonic control approach [2] in which a feasible solution emerges from the combination of individual material handling holons’ solutions. Leitão and Restivo [11] presented a holonic approach to manufacturing scheduling, where the scheduling functions are distributed by several entities, combining their calculation power and local optimization capability. Despite the aforementioned results, there is still a lack of study on process planning in HMS.

Planning of processes in HMS is driven by customers’ orders. Given an order with specific product demands, the holonic process planning problem is to find and organize a set of holons in HMS to compose production processes to
fulfill the order requirements. The objectives of this paper are to propose a methodology to develop a problem solver for the holonic process planning problem. Contract net protocol (CNP) [15] is a well known protocol for distributing tasks. Application of CNP for task allocation in HMS is found in [13], [5], [14]. Formation of holonic processes in HMS based on CNP has been studied in [6], [7], [8], where Petri net [12] models have been proposed to capture the interactions between resource holons and product holons. These results pave the way for the development of a problem solver for composing holonic processes to meet the order requirements. To achieve the research objectives, we combine multi-agent systems technology with Petri nets in this paper to design and implement HMS to fulfill the requirements of orders. We first propose architecture to describe the negotiation between order holons, product holons and resource holons in HMS and formulate holonic process planning problem based on Petri nets and propose a method to find an optimal solution. To determine whether it is feasible for a set product holons and resource holons to meet the requirements of an order holon, we propose Petri net models, which can be specified in Petri Net Markup Language (PNML) [18], to capture the workflows and activities in product holons and resource holons, respectively. To enable a holon to discover the services provided by other holons, we propose a scheme for publication and discovery of holon services based on FIPA compliant multi-agent system platform. Based on the service publication and discovery scheme, the two-layer contract net protocol and the Petri net models of holons, we develop a problem solver to solve the holonic process planning problem.

The remainder of this paper is organized as follows. In Section 2, we describe and formulate the holonic processes planning problem (HPPP). In Section 3, we propose a two-layer contract net protocol for solving HPPP. In Section 4, we present the platform for publication and discovery of holons. In Section 5, we first detail our design using UML activity diagrams to describe handling of the two layer protocol messages and then demonstrate our implementation using an example. Section 6 concludes this paper.

2. Holonic System Operation

An HMS consists of three types of holons: resource holons, product holons and order holons [3]. A resource holon consists of production resources with relevant components to control the resources. A product holon contains the production process information to manufacture products. An order holon represents an order. Individual product holons or resource holons cannot process a complex task alone. Fig. 1 shows our proposed architecture for processes planning in HMS. In Fig. 1, the workflow model repository is used to capture the workflows of product holons. The resource activity model repository is used to capture the activities of resource holons. The HMS Designer is the core of our proposed system to model and construct the holons in HMS, activate negotiation of holons, monitor the system operation and keep track of the established contracts. Depending on the role of a user involved in the system, the accessible functions and user interfaces vary. An administrator is responsible for managing the account of users. An order manager is responsible for creation of an order holon and initialization of the two-layer contract net protocol. A product manager is in charge of the defining and creating product holons. A resource manager may define and update resource holons. The HMS Designer aims to provide a multi-agent problem solving platform and negotiation mechanism to coordinate holons and monitor the progress of tasks required to fulfill the order requirements.
To process an order, a set of resource holons and product holons form a composite holon called a holarchy. The problem to determine the best product holons and resource holons to fulfill the order based on the available resources can be formulated as an optimization problem as follows. Let \( c(H, R) \) denote the holarchy formed by a set \( H \) of product holons and a set \( R \) of resource holons. Let \( w: \{c(H, R)\} \rightarrow R^+ \) be the cost function that specifies the cost of \( c(H, R) \), where \( R^+ \) denotes the set of positive real values. The following problem is formulated to minimize the cost while meeting the time constraint.

Holonic Process Planning Problem (HPPP): \( \min w \) subject to the constraint that \( c(H, R) \) is feasible and satisfies the time constraint, \( R \subseteq R, H \subseteq H \), where \( R \) denotes the set of all potential resource holons and \( H \) denotes the set of all potential product holons.

One way to form a holarchy is based on the contract net protocol (CNP). Interactions among resource holons, product holons and order holons are through the well-known contract net protocol [15]. In contract net protocol, there are two types of roles an agent can play: manager or bidder. Four stages are involved to establish a contract between a manager and one or more bidders: (1) Request for tender (2) Submission of proposals (3) Awarding of contract and (4) Establishment of contract. CNP can be applied to form a holarchy base on establishment of contracts between a set of product holon and a set of resource holons.

3. Two Layer Contract Net Protocol

HMS provides a flexible architecture to configure product holons to fulfill the order requirements. We propose a two-layer contract net protocol that consists of an upper layer protocol and a lower layer protocol to solve HPPP. The upper layer contract net protocol is applied by an order holon and the potential product holons to find the minimal cost product holons that form a holarchy. After conducting the upper layer protocol, a holarchy is formed by a set of product holons. To acquire the resources required to execute the operations in the holarchy, the lower layer contract net protocol is applied between the holarchy and the potential resource holons to analyze the feasibility and determine the resource holons. Resource holons, product holons, and order holons are assigned different roles at different phases in the two-layer protocol. To describe the negotiation processes between an order holon, the product holons and the resource holons in HMS, the following messages are defined.

\[ h_1 \quad h_2 \quad h_3 \quad h_4 \quad h_5 \quad h_6 \quad h_7 \quad h_8 \]

\[ x_{\text{rft}} \quad x_{\text{sop}} \quad x_{\text{aoc}} \quad x_{\text{eoc}} \]

Fig. 2 (a) Request for tender  (b) Submission of proposals   (c) Awarding of contracts      (d) Establishment of contracts  (e) A holarchy

- \( h_1 \): Order holon
- \( h_2 \): Product holon
- \( h_3 \): Resource holon

Definition 3.1: \( x_{\text{rft}} \) denotes a “Request for tenders” message sent by either an order holon or a product holon. \( x_{\text{sop}} \) denotes the “Submission of proposals” message sent by either a product holon or a resource holon. \( x_{\text{aoc}} \) denotes the “Awarding of contract” message sent by an order holon or a product holon. \( x_{\text{eoc}} \) denotes the “Establishment of contract” message sent by an awarded product holon or an awarded resource holon.

The upper layer protocol specifies the interactions between an order holon and the potential product holons. The order holons always act as managers to initiate tender processes with the product holons as the bidders. The product holons act as the managers in turn to issue a new request for bids to other product holons for the set of subtasks stemming from the product requirements.
Example: Figure 2(a)–(d) illustrate a scenario in which a production process is to be formed in HMS based on the upper layer contract net protocol applied between the potential product holons \( h_1, h_2, \ldots, h_7 \) to fulfill the requirements of an order holon \( h \). Holonic processes are production processes dynamically created based on the collaboration of product holons. Each product holon has an internal process flow, the required input types and output types. Figure 2(e) shows the holarchy \( H \) formed resulting from the upper layer protocol. The cost of a product holon is obtained based on the costs of the operations.

The lower layer specifies the interactions between the holarchy formed by combining the product holons obtained from the upper layer protocol and the potential resource holons to analyze the feasibility of the solution. Figure 3(a)–(d) illustrate a scenario in which the lower layer contract net protocol is applied between the holarchy \( H \) and resource holons \( r_1, r_2, \ldots, r_6 \).

A timed Petri net (PN) \( G \) is a five-tuple \( G = (P, T, F, m_0, \mu) \), where \( P \) is a finite set of places with cardinality \( |P| \). \( T \) is a finite set of transitions, \( F \subseteq (P \times T) \cup (T \times P) \) is the flow relation, \( m_0 : P \rightarrow \mathbb{Z}^{|P|} \) is the initial marking of the PN with \( Z \) as the set of nonnegative integers and \( \mu : T \rightarrow \mathbb{R}^+ \) is a mapping that specifies the firing time for each transition. The marking of \( G \) is a vector \( m \in \mathbb{Z}^{|P|} \) that indicates the number of tokens in each place and is a state of the system. A transition \( t \) is enabled and can be fired under a marking \( m \) if and only if \( m(p) \geq F(p,t) \). Firing a transition once removes one token from each of its input places and adds one token to each of its output places. A marking \( m' \) is reachable from \( m \) iff there exists a firing sequence bringing \( m \) to \( m' \). The readers may refer to [12] for a tutorial on Petri nets. Details on how to construct Petri net workflow models and resource activity models can be found in [8]. In modeling a holon with Petri net, we use a place to denote a production state whereas a transition to represent an operation. The workflow Petri net of a holon \( h \) is defined as follows.

Definition 3.2: The workflow Petri net of product holon \( h \) is an acyclic Petri net \( W_h = (P_h, \cdot P_h \cup P_h^{in} \cup P_h^{out}, T_h, F_h, m_0, \mu_h) \), where each place in \( P_h \) has exactly one input transition and exactly one output transition, each place in \( P_h^{in} \) has exactly one output transition but no input transition and each place in \( P_h^{out} \) has exactly one input transition but no output transition and \( m_0(p) = 0 \) for all \( p \in P_h \).

Fig. 3(e) shows an example of workflow Petri net.

Resource activity Petri net models are proposed to capture the dynamics of resources. Let \( R \) denote the set of resource types in the system. A type-\( r \) resource may take part in a set of activities \( \Omega_{hr} \) involved in \( W_h \).

Definition 3.3: The \( k-th \) activity of type-\( r \) resources, where \( k \in \Omega_{hr} \), is described by an activity Petri net \( A_r^k = (P_r^k, T_r^k, F_r^k, m_{r0}^{in}, \mu_h) \), where \( m_{r0}^{in}(p) \) is the number of resources allocated.

Fig. 3(f) shows three resource activity models. Remark that \( T_r^k \cap T_{r'}^{k'} = \emptyset \) for \( k \neq k' \). If a type-\( r \) resource is not used, it stays at idle state place \( p_r \). To represent the idle states for all types of resources, we define a set of idle state
places as follows. Fig. 3(h) illustrates the proposed scheme to construct the Petri net model for feasibility analysis. Construction of the Petri net model is obtained by taking into account the interactions between the resource activity and the relevant workflow of the task. A composition procedure is needed to combine the resource activity and the task workflow into a process model. Based on the process model, the feasibility analysis is performed.

4. PUBLICATION AND DISCOVERY OF HOLONS

We develop a system based on Java Agent Development Environment (JADE) to realize our methodology. JADE is a middleware that facilitates the development of multi-agent systems. Messages exchanged by JADE agents have a format specified by the ACL language defined by the FIPA international standard for agent interoperability. To apply the contract net protocol, a manager must be able to search for the services provided by the other bidders. Each time a holon is added to the system, its services are published through the DF (Directory Facilitator) agent in JADE. JADE provides a ServiceDescription class to specify the serviceName, serviceType and add new properties for the services provided. In our prototype, we define different service types for order holons, product holons and resource holons. Each type of holons can be implemented as an agent in JADE platform. JADE platform manages agents with AMS (Agent Management System). The two-layer contract net protocol relies on an infrastructure for individual holons to publish and discover their services. The JADE platform provides a built-in directory service to simplify publishing and discovery of services. An agent wishing to publish one or more services must provide the DFagent with a description including its AID, possibly the list of languages and ontologies that other agents need to know to interact with it and the list of published services.

![Diagram of publishing services, discovering services, request for tender, and discovering services]

Fig. 4(a) shows a scenario in which holons $h_1$ ~ $h_8$ publishing their services via DF agent to provide yellow page services to other holons. For each published service a description is provided including the service type, the service name, the languages and ontologies required to exploit that service and a number of servicespecific properties. In order to publish a service an agent must create a proper description (as an instance of the DFAgentDescription class) and call the register() static method of the DFServic class. When an agent terminates it is a good practice to de-register published services. To discover the services provided by the potential bidders, a holon must search via the yellow page services. A holon searching for services must provide the DF with a template description. The result of the search is the list of all the descriptions that match the provided template. A description matches the template if all the fields specified in the template are present in the description with the same values. Fig. 4(b) shows how order holon $h_8$ searches the yellow page for the potential product holons. Suppose order holon $h_8$ discovers product holons $h_7$ that can provide the requested services. Order holon $h_8$ will send a “Request for tender” message to product holon $h_7$ as shown in Fig. 4(c). On receiving the “Request for tender” message, product holon $h_7$ searches the yellow page to discover the potential bidders as shown in Fig. 4(d).

5. HANDLING PROTOCOL MESSAGES

To realize the proposed idea, we implement a holonic system based on the two-layer contract net protocol. In this section, we first present the activity diagrams for handling the two-layer contract net protocol messages and then illustrate the developed system. The activity diagram for handling a “Request for tender” message $X_{rft}$ is shown in Fig.
5(a). In handling a $X_{rf}$ message, a holon first checks whether the product type is supported. If it is supported, the holon checks whether the timing constraint is satisfied based on the Petri net model. The method proposed in [6] is adopted in this paper to check whether the timing constraint can be satisfied. Fig. 5(b) shows how to handle “Submission of proposal” messages $X_{sop}$ from the bidders. Fig. 5(c) shows the activity diagram for handling a “Request for tender” message in the lower layer. The host product holon determines the winners by conducting feasibility analysis based on Petri net model ([8]). Please refer to [8] for the details of the algorithm to conduct feasibility analysis based on Petri net model. Figure 5(d) shows how to handle “Submission of proposal” messages $X_{sop}$ in the lower layer contract net protocol.

![Diagram of handling upper layer $X_{rf}$ message](image)

![Diagram of handling upper layer $X_{sop}$ message](image)

![Diagram of handling lower layer $X_{rf}$ message](image)

We develop a HMS Designer based on the two-layer contract net protocol proposed in this paper. We illustrate the function of HMS Designer by applying it to an application scenario. HMS Designer is developed based on the Petri net models of holons, the services publication/discovery mechanism in Section 4 and the message handling activity diagrams proposed in Section 5. Fig. 6~Fig. 9 shows the screen shots of our system, which consists of a HMS Editor/Monitor (Fig. 6) and graphical user interface to set the properties of order holons, product holons and resource holons (Fig. 7~Fig. 9) and display the established contracts between order holons, product holons and resource holons based on our proposed two-layer contract net protocol. The following example details an application scenario of the HMS Designer.

Example: Consider the requirements of an order with requested product type: Product Type 5, order due date: 30, and cost constraint: 80.

Fig.6 demonstrates the HMS Editor in which users (order holon managers, product holon managers and resource holon managers) may define the holons in a system. User may drag and drop the order holon, product holon and resource holon icons to define the holons. For this example, there are thirteen holons ($h_1 \sim h_{13}$) that have been defined, including seven product holons and one order holon. Fig. 7 shows the graphical user interface for order holon managers to set order holon properties. The properties of an order holon include holon id, model (represented by Petri net in PNML format), input/output types and costs, where input product types is the product types required for producing the output product types.

In Fig. 7, the product type, due date and cost of order holon $h_k$ is set to Product Type 5, 5, 30 and 80, respectively. Fig. 8 illustrates the graphical user interface for product holon managers to set product holon properties. The properties of a product holon include holon id, model (represented by Petri net in PNML format), input/output product types and costs, where input product types is the product types required for producing the output product types.

Fig. 9 illustrates the graphical user interface for resource holon managers to set resource holon properties. The contracts between holons are represented by directed arcs between holons as shown in Fig. 6.
6. CONCLUSION

To meet the changing customers’ demands with minimal costs while delivering the products timely, manufacturers rely on a flexible architecture to dynamically collaborate with the partners to fulfill the orders’ requirements, optimize the costs and reduce the risk. Holonic manufacturing systems provide a reconfigurable, flexible and decentralized manufacturing environment to accommodate changes and meet customers’ requirements dynamically based on the notion of holons. This paper presents a systematic methodology to synthesize minimal cost production processes that satisfy timing constraints based on holonic system architecture. A holonic process is dynamically formed by a set of product holons and resource holons to execute a task. We consider the problem to synthesize holonic processes with minimal costs while meeting the timing constraints in holonic systems. We formulate this problem based on a hybrid model in which contract net protocol is adopted as the negotiation protocol and Petri net is used to specify and analyze the timing and resource constraints. We then propose a two layer contract net protocol to describe the negotiation between order holons, product holons and resource holons in HMS. To determine whether it is feasible for a set product holons and resource holons to meet the order requirements, we introduce Petri net models to capture the
workflows and activities of product holons and resource holons, respectively. We propose a scheme for publication and discovery of holon services based on a FIPA compliant multi-agent system platform. Based on the service publication and discovery scheme, we develop a problem solver for holonic process planning. We demonstrate the holonic process planning solver by an example. As our methodology is generic and developed based on holonic system architecture, it can be tailored for specific HMS applications.

ACKNOWLEDGEMENTS

This paper is supported in part by National Science Council under Grant NSC97-2410-H-324-017-MY3.

REFERENCES


Leveraging new SEMI standard to reduce waste and improve flow

Raymond Goss¹, Carmen Maxim, Diwas Adhikari, Jan Rothe
GLOBALFOUNDRIES
5113 Southwest Parkway, Suite 300, Austin, TX 78735, USA

ABSTRACT
A successful cutting-edge semiconductor manufacturer applies Lean principles with the goal set on perfecting its operation flow. In 2007 SEMI adopted a new standard E94-1107 that breaks the dependency between carrier and lot for processing wafers and formally introduces material redirection, which creates several opportunities for waste reductions and improved WIP handling. As depicted from simulations, the new standard is impacting several KPIs including throughput, cycle time, yield and Just-In-Time customer response. This paper discusses different lean optimizations leveraging the new standard. Some tools require loading of many wafers simultaneously to achieve a high throughput. Although with reducing lot sizes, takt time can be improved, the sequential nature of these tools become limited by the number of load ports and thus have a negative impact on the equipment effectiveness. Before the new standard implementation, no carrier could be removed from load ports during processing to allow other carriers to load or unload wafers leaving the ports blocked. With this change, we can now remove the carrier from the Load Port after unloading wafers into the tool to allow filling the tool with the optimum number of wafers. Over the years, numerous lean manufacturing studies have been performed in terms of determining the correct lot size (number of wafers in a lot) to meet the Just-In-Time targets. Simulation runs demonstrate that smaller lots have up to 33% and 50% improved cycle time for batch tools and single wafer tools respectively. Based on fluctuations in demand, there may not only be a need to speed up, but also to slow down production quickly. All these have direct correlation with scheduling of right sized lots (dynamic lot sizing), carefully planned delivery of carriers for processing and improved pull production techniques.

1. INTRODUCTION

Lean Manufacturing is an operational strategy that focuses only on value added activities and eliminating waste. It has its origins from the Toyota Production System and its main goal is to increase the value-added work. Value is defined as an item or feature for which a customer is willing to pay. All other aspects of the manufacturing process are deemed waste. In the semiconductor manufacturing space there are several types non-value added activities that can be mapped to the initial seven wastes:

- Produce more wafers than needed to compensate for low yield; The production unit is wafer as opposed to die/product that is requested by a customer (Overproduction)
- Lot waiting without being processed (Waiting)
- Redundant moves, poorly scheduled deliveries (Transportation)
- Eliminate white spaces (Processing times, Waiting)
- Unfinished products, excess test wafers (Inventory, Overproduction)
- Yield problems, scraped wafers (Defects)

Lean manufacturing is used as a tool to focus resources and energies on producing the value-added features while identifying and eliminating non value added activities. The benefits generally are lower costs, higher quality, and shorter lead times.

¹ Corresponding author: E-mail: Raymond.Goss@globalfoundries.com
2. **JUST IN TIME**

Just-In-Time (JIT) is the backbone of the lean manufacturing. Just-In-Time philosophy is that resources are made available only in the amount required and at the time required. Small lot production (ideally one piece) is an important component of many JIT strategies. Lot size directly affects inventory and scheduling. Figure 1 shows a comparison between Large and Small lots. JIT produces the same amount if setup is the same. Small lots also increase flexibility to meet customer demand.

![Small Lot vs. Large Lots processing](image)

Figure 1: Small Lot vs. Large Lots processing

### 2.1 Semiconductor Industry Background

Semiconductor industry is more and more concerned with continuous improvement and reducing waste. Lean initiatives are ongoing in many areas including design, development, testing, deployment and manufacturing.

Semiconductor device fabrication is the process used to create integrated circuits. It is a continuous flow of multiple-step sequence of processing steps (coat, expose, develop) during which electronic circuits are gradually created on a wafer made of semiconducting material. The entire manufacturing process from start to packaged integrated circuits ready for shipment takes six to eight weeks and is performed in highly specialized facilities.

A single fabrication facility (fab) in the semiconductor industry costs billions of dollars and is becoming increasingly more expensive. The process and tool costs increase with each new technology size shrink. These technology node changes are inevitable to keep up with Moore’s law ([5]). The change between 130nm to 90 nm and to 65nm took 2 years each ([6]). While these changes may require more costly tools modifications and may add processing steps, the increase of wafer size dictates complete replacement of all hardware including transportation systems. Although GLOBALFOUNDRIES completely replaced all 200mm tools to convert Fab 30 in Dresden Germany into 300mm Fab 38 which also required ceiling reinforcements to support a new transportation system, conversions are much more costly and occur much less frequently. The semiconductor industry is debating over the when the next wafer size migration to 450mm will need to take place. Many, including GLOBALFOUNDRIES, believe that we can achieve higher efficiencies and benefits with the existing platforms, which will help contain budget expenses especially when times are as financially difficult as they are now. This lean initiative is designated as 300mm Prime.

3. **E94 Description**

Success of the 300mm Prime transformation will be verified by improving 17 Key Performance Indicators ([2]). SEMI defines standards for all equipment and processing for the semiconductor industry. One very promising SEMI change was approved late 2007 (E94-1107) which directly enables 300mm Prime optimizations and targets key bottleneck operations. This change relaxes the strict relationship between the equipment Control Job and the carrier. Prior to this modification, the source and destination carriers and wafer maps were fixed at the start of the process operation and could not be changed during processing. The usual implementation required that the input and output specification to be identical. This new SEMI standard version now supports material redirection and allows a late determination of the carrier and wafer map up until the first wafer completes and needs to move out of the equipment.

Figure 2 shows how the change removes the limitations. The source carrier can now be removed from the equipment load port (LP) after all the wafers to be processed are moved into the equipment. The target carrier for the wafers of a job can be specified after the processing has started but before the first wafer finishes processing.
Leveraging new SEMI standard to reduce waste and improve flow

Once the factory systems have identified a proper destination carrier (before the first wafer of the control job finishes processing), the destination carrier is transferred to the equipment LP just-in-time to receive the first wafer after it completes processing. The equipment software updates the factory tracking information with the destination carrier and the processed wafers move into the carrier. When all the wafers have finished processing and have moved to their destination, the carrier will be ready to be transferred to a storage location or the next equipment. Note that using the source carrier of a previous lot that just finished delivering its wafers and that could be otherwise be unloaded from the equipment, as a destination carrier could also be leveraged in order to save several transportation steps.

4. Dynamic Lot Sizing

Applying the lean principles usually involves looking at a typical triangle cost-quality-delivery balance. An often overlooked component of this is quantity. If we focus on our customer (perceived quality), we can add quantity to the triangle. There is a paradigm shift taking place with applying lean to semiconductor industry in regards to the impact of quantity on delivery, costs, and value provided to the customer. With 200mm technology, wafers were typically processed in lots of 25 with the only differentiator in the lot being the slot in which a wafer was. Most Fabs are filled to maximum volume to achieve low costs per wafer and maximize tool utilization. With the introduction of 300mm, wafers can be tracked at the wafer level but are still mainly tracked and moved in lot groups. There is promise of optimizing equipment throughput by grouping wafers in lots smaller than 25. While it requires more transports to move the same number of wafers if the lot size is smaller, smaller lot sizes provide the opportunity to move lots faster through some bottleneck equipment. This allows a greater mix of products and allows priority “rocket” lots to be more efficiently scheduled and accelerated through the fab. Also, not all products at all stages of development require large volumes of production. Providing customers with quicker deliveries can actually lead to less overproduction, higher value on what is produced, and corresponding higher average selling prices. The corresponding WIP inventory is reduced. This actually means less volume can be more profitable.

4.1 Small Lot Manufacturing

Figure 2: New SEMI Standard description

Figure 3: Cycle time reduction – Batch Tools simulation vs Single Wafer Tools simulation
In recent years, numerous lean semiconductor manufacturing studies have been performed to determine the correct lot size (number of wafers in a lot) to meet the JIT targets. Figure 3 left outlines the simulation results for the relation between lot size and cycle time improvements in a factory using batch processing, while right depicts the results for a simulated fab using single wafer processing only. However, shifting to smaller lot sizes or Small Lot Manufacturing (SLM) adds complexity to other components of Factory Automation systems.

4.2 A SMALL LOT CHALLENGE EXAMPLE

For the sake of simplicity we will assume a lot size of 12 wafers and consider wafers processing at particularly critical bottleneck equipment, the Advanced Lithography Cell (ALC), configured with 5 load ports (LP) and a capacity to simultaneously process 70 wafers. It would be restricted to a maximum of 60 wafers if a carrier only contained 12 wafers, which would lead to a reduction in equipment throughput for standard scenarios. Allowing temporary removal of the carrier from the equipment LP once the wafers have been moved into the tool would solve this problem. This scenario could also support redirection of completed wafers into a different carrier (the carrier that delivers the wafers may be different from the one that comes to pick them up). The maximum benefit is actually achieved by supplementing the redirection capability and reusing the newly emptied carriers to pickup completed wafers. This creates a carrier cascade (dropping off and picking up wafers) thus maximizing throughput and LP optimization.

The new SEMI specification provides the solution for this fab scenario. The ability to redirect completed wafers into a different carrier means the source carrier no longer has to remain clamped to the equipment LP, thus freeing up the LP for the delivery of additional wafer lots. This not only enables the ALC to run with its optimal capacity of 70 wafers, the redirection and cascading support eliminates wait time for empty carrier and results in high equipment efficiency. Working with major lithography track suppliers, our company has already qualified the implementation of the material redirection mode in the equipment software of one of these vendors for production use.

4.3 ULTRA SMALL LOT MANUFACTURING

Further cycle time reduction can be attained by reducing the lot to one or two wafers (ultra small lot). It has been proposed by GLOBALFOUNDRIES and others in the industry ([1], [4]) that single wafer processing can improve cycle times by 44 percent on batch tools and 76 percent on single wafer processing tools, which ultimately provide faster product turnaround to the customer [1]. An example of a maximum possible benefit can be seen in . Since wafers on single wafer processing equipment are independent from other wafers within the same lot to complete processing the cycle time to process one lot can be significantly reduced.

During the whole manufacturing cycle wafers are spending a significant amount of time (30-50%) [1] in the stockers. Between process steps it is expected that a lot with 1-2 wafers to go more frequently to the next tool rather than being returned to the stocker. Each move to the stocker adds cycle time and complexity. However, in the USLM paradigm move moves will be performed. As a consequence the new automated material handling system should be designed to the anticipated 10x more moves per hour for an equivalent fab throughput.

Moreover, the ultra small lots have a potential of achieving “one piece flow” for semiconductor manufacturing. Special adaptations need to be made to the hardware and software. Running a fab in mixed mode (regular and Ultra Small Lots) or only with ultra small lots requires several changes to the equipment, transport and factory operation modes. Some of the challenges are described in the next section.

5. SINGLE TRANSPORT

A single wafer lot comes with its own set of challenges for equipment having internal buffers, batch processing capability or any ability to process more than one wafer at a time. Automated Material Handling Systems needs to be adjusted and be able to handle small lot size carriers which can carry one or two wafers. The solution of a USL carrier will make possible a reduction in cycle time and improve efficiency of wafer storage.

Running with small lots had the potential to reduce wafer contamination. The carrier door will be less opened and thus limit the amount of exposure. This feature will be useful in case for instance of wafer sampling. In a typical case of sampling 10 wafers from a lot with containing 25 wafers will expose to contamination the rest of 15. This could lead to an increase in wafer yield. Stockers need to be redesigned as well. The shelf spacing should be reduced or made flexible. Some viable alternative could be
Leveraging new SEMI standard to reduce waste and improve flow

Bare Wafer Stocker that stores wafers in stacks (of 100 for example) and have sorter capabilities. The new types of carriers have been improved with a horizontal magnetic door that will have less mechanical failures.

Figure 4: Modified load port configuration for 3 ultra small lot FOUPs instead of one 300mm FOUP – Brooks Automation, Inc.

Changes to load port configurations need to be done to prevent equipment starvation for tools that requires more wafers than load ports.

Figure 5: Ultra SLM Shelves from Advanced Development Center of GLOBALFOUNDRIES
6. TEST WAFERS

Quality in a fab is maintained through continuous improvements identified via real time monitoring of the end to end processes. Highly integrated applications capable of making complex real time decisions are monitoring every aspect of the fabrication process. Any deviation can be corrected real time to maintain quality and to minimize waste. For example, results from pre- and post-process metrology step during process monitoring can be used to automatically adjust recipe parameters settings in both upstream and downstream processes. If wafers get a thicker coat during deposition, the downstream polish recipe parameter can be automatically adjusted to polish more. Similarly, the information can be fed upstream such that recipe parameter gets automatically adjusted for deposition to apply a thinner coating. Equipment in a fab is routinely calibrated to ensure their process consistency, as preventive maintenance (scheduled based on time or usage), and after unexpected failures.

A special category of wafers called test wafers (TWs) are used for process monitoring and tool calibration. Some process inspections are intrusive in nature and “destroy” some wafers. A raw wafer typically cost around 1K, and increases in value as its gets processed on the line amounting to thousands of dollars. For this reason product wafers are rarely used for testing and monitoring purposes. Instead huge inventories of lower value TWs are maintained. A fab with 25K production wafer starts per month may easily have twice as many TWs in its inventory.

TWs require a complex life cycle management. The logic to determine the wafers of correct grades for process monitoring or equipment calibration purpose is cumbersome – whether it is performed manually or automatically. The Figure 6 depicts a TW management loop that encompasses a fairly simple calibration flow:

- TWs usually start spread out in multiple carriers (stored in shelves or stockers)
- When TWs are required for either processing monitoring or calibration:
  - Correct grade of TWs are selected
  - TWs get consolidated into a single wafer carrier using a sorter
- TWs are prepped for specific process monitoring or calibration. This might require several operations on a flow.
- Upon completion of the monitoring or calibration activities, TWs may be:
  - Reused – TWs are immediately available as usable inventory
  - Recycled - TWs get downgraded and/or cleaning activities maybe required.
Leveraging new SEMI standard to reduce waste and improve flow

- Scrapped – TWs are no longer usable.

The sheer volume of TWs, the complexity of their management, and need for smart and timely selection of correct grade wafers for process monitoring and calibration makes it a prime candidate to apply lean principles. A lean focus on TWs has the potential to reduce TW inventory and prevent product inventory pile up by reducing wait times for equipment.

6.1 SMART INVENTORY MANAGEMENT

The reclaim, recycle, downgrade and reuse path for the TWs must be clearly defined so that we get the most out of a given test wafer before it gets scraped. To ensure that a TW is optimally used the evaluation for downgrade, reuse, and other prepping related to TW management must be well defined and executed. Sharing of TWs across multiple modules (sections) of the manufacturing in the fab could result in higher efficiency resulting in reduced inventory and storage requirements.

6.2 SMART SCHEDULING

During process monitoring, TWs are pre-measured, “batched” with the actual product wafers, processed, and finally post-measured. If production wafers wait due to unavailable process monitoring TWs production inventory build-up may occur. With smart scheduling we can ensure JIT TW availability to be batched with the production wafers. Ideally, the post-measurement activities should be accomplished as quick as possible to provide results on the production line to reduce cycle time or adjust process control. TWs scheduling for calibration should anticipate preventive maintenance (scheduled, or usage base) and have them available JIT. Poor TW scheduling would otherwise lead to unnecessary slow-downs or require the acquisition of more TWs.

6.3 SMART TOOL USAGE

Equipment is limited by the number of ports. One port can only hold one carrier at a time. Furthermore, once a port is occupied, the carrier cannot be removed until the wafers are returned from the equipment. In the TW scenario, some of the pre- and post-measurements are non-sequential operations. It is usual for different wafers in the same carrier to be measured in different measurement operations. However, due to the current restrictions with tool port limitations and lack of material redirection support, all the wafers are first consolidated into a single carrier and sequentially measured at the pre- and post-measurement operations. This has the following negative impacts:

- The additional sorting steps take time and require enough sorting tools.
- TWs in the same carrier go through needless waiting on measurement tools even when they are not involved in the measurement.

Figure 7, illustrates a use case for a calibration flow requiring TWs W1, W2, W3 and W4:

Figure 7: TWs Flow without Redirection Support

- A sorting task is first executed to consolidate the required wafers into the single carrier.
- In this example, TWs W3 and W4 wait in the carrier while W1 and W2 are measured in PreMeas1 and likewise W1 and W2 wait at PreMeas2.
Similar waiting occurs again at the post-measurement operations.

Leveraging the new SEMI standard, if the factory systems and the tool vendors support the redirection feature and introduce additional equipment buffer storage, facilitates a leaner process, eliminating the sorting task and increasing the overall efficiency of the TW flow, as seen in Figure 8.

In cases where the pre- and post-measurements can be performed non-sequentially, the carriers containing the TWs can be directly dispatched to the metrology equipment. Figure 8, illustrates a redirection enabled calibration use case that still requires that all wafers to be in a single carrier at the processing operation.

- The carriers containing W1, W2 and W3, W4 respectively are simultaneously dispatched. The carriers then can either:
  - Remain on the equipment port waiting on completion of the measurement activity.
  - Pick up any measurement completed wafers from previous run and proceed to its next destination.
  - Simply proceed to its next destination.
- When pre-measurements nears completion, a carrier can be scheduled (and dispatched if required) to pick up the TWs from the metrology equipment, and then dispatched to the processing operation.
- On completion of the processing step, the TWs can then be dropped off to their respective post-measurement equipment.

When the redirection is enabled for metrology equipment, there is a possibility to optimize TW management.

7. SUMMARY

In 2007 SEMI adopted a new standard E94-1107 that breaks the dependency between carrier and lot for processing wafers and formally introduces material redirection, which creates several opportunities for waste reductions and improved WIP handling. As depicted from simulations, the new standard is impacting several KPIs including throughput, cycle time, yield and Just-In-Time customer response. This paper discusses different lean optimizations leveraging the new standard.

The new SEMI standard offers the capabilities to apply lean principles across wafer manufacturing. Some of the results afforded are: reduced inventory, parallel processing to minimize wait times, optimize use of carriers and deliveries and elimination of non-value operations.

Finally, with the ability to process individual wafer lots, “single piece flow” is achievable in the semiconductor industry.

ACKNOWLEDGEMENTS

Our colleagues at GLOBALFOUNDRIES have been instrumental in providing us data and support. We especially recognize Achim Felber, Gerald Goff, Jeff Hanan, Gilbert Vigil, and Olaf Zimmerhackl. We also thank Entegris, Brooks Automation Inc., and Murata for allowing the use of images of their products (Figures 4 and 5).
Leveraging new SEMI standard to reduce waste and improve flow

REFERENCES


ABSTRACT

Lean production comprises a set of different tools geared towards the elimination of all operations that do not add value to a product, service or process, thereby increasing the value of each activity and removing all that is not required. An essential part of lean production is training and education. Training and educating the workforce (what we refer here as “soft investment”) will create the necessary conditions to engage and involve employees in improvement activities, so are indispensable in the implementation of lean production. The aim of this study was to explore and contrast the relationship between “soft investments” in lean firms and “soft investments” in traditional firms, by set of hypotheses tests. Data, both quantitative and qualitative, were collected through a survey in conjunction with short structured interviews and planned visits to several organisations. Two questionnaires were administered at two different levels of the organisations: one to the CEO or general managers and the other to the operations managers with the objective of evaluating the relationship between “soft investment” and the implementation of lean production. More than 30 firms in the tableware industry of the UK completed the questionnaires. The information collected allowed an analysis of the given hypotheses and the results and conclusions of the study shown that organisations that use lean manufacturing invest significantly more in training the labour force than those that do not.

1. INTRODUCTION

This study investigated the application of lean production techniques in the UK ceramics industry of North Staffordshire. This craft-based industry represents a major manufacturing sector in the UK and it depends heavily on export markets, the majority of which are composed of tableware products. Although the sector is still a major employer in North Staffordshire, it is estimated that more than 44 per cent of jobs have been lost in the last 25 years [1-6]. Closures, downsizing and restructuring have taken place in many ceramic firms, mainly in response to changes in market demands, increasing foreign competition and the introduction of new technology.

Lean manufacturing pursues simultaneous improvements in all aspects that drive the elimination of waste. This is usually achieved through projects that change the physical organisation of work, logistical and production control throughout the entire supply chain; the way the human effort is applied not only in production activity but also in the support activities. Lean manufacturing was developed first in Japan by the instigators of the Toyota production system: Taiichi Ohno, Shingo Shingo, Eijy Toyoda among others. The lean manufacturing system has been defined as a philosophy founded on different principles, see [7].

* Corresponding author: Tel.: +44(0)1604892185; E-mail: Horacio.SorianoMeier@northampton.ac.uk
Researchers from MIT, theoreticians of the new manufacturing paradigm called lean production, conducted an international investigation of the automotive industry and differences in productivity entitled the International Motor Vehicle Program (IMVP), costing $5M. The MIT academics argued that this system would change the organization of production, and that its adoption was absolutely necessary in US and European industries to remain competitive in world markets.

Since the publication of the lean production thesis [8], the interest in this concept has grown and evolved considerably. The diffusion of information regarding the lean concept can be obtained in key books and articles, which are often quoted by academics since the lean movement was founded. Among these references it is important to mention the contribution of Womack’s books: The Machine that Changed the World (1990); Lean Thinking: Banish Waste and Create Wealth in your Corporation (2003); Lean Solutions: How Companies and Customers Can Create Value and Wealth Together (2005). Other key books include those published by Shigeo Shingo [9] and Taiichi Ohno [10], and the paper published by Spears and Bowen [11]. The lean concept has many appeals for the practitioner via just-in-time management, and the integration via computer-aided processes to the areas of design, factory management, supply and distribution, see [12]. Womack and Jones [13] argue that lean production can be applied to any organisation or economic activity. They also argue that lean production provides a way of generating sources of employment instead of destroying the same on behalf of efficiency and that managers have turned ‘mean’ instead of ‘lean’ as a consequence of not employing the system correctly or simply employing it piecemeal. Lean production is a long term approach and not a panacea to sort out competitive problems in the short run [8]. Therefore, lean manufacturing is best viewed strategically as a formidable weapon in increasingly competitive markets [14]. Theoretically and critically lean production also appeals to academics. It represents a natural progression from Fordist mass production, although there has also been a debate on the extent to which it represents a new paradigm [15], [16]. After years of ‘downsizing’ and ‘re-engineering’, managers needed to look for a model that would allow them to engender continuous and successful growth.

2. EDUCATION AND TRAINING – THE LEAN CONTEXT

The elements of lean production are evident through different industrial sectors, but the rate of change is dramatically different and the specific results can vary from firm to firm [17]. Despite this interest, nonetheless, we really do not know the boundaries of this lean production system. Therefore, there is a need to conduct more research in this important area. Since the introduction of this production system [8] this has been the topic of more research in operations management. In terms of its implementation, McClure [18] indicates that lean production may find a subtle form of resistance. However, this could be prevailed over by the training of employees in preliminary steps. Grieves [19] argues that people have a dominant role in the implementation of lean production, stressing the need of training and educating the workforce. In addition, Slack et al [20] argue that, under a lean perspective, training is a key factor when improving and implementing processes.

Bigelow [21] defines training as “formal instructions sessions to provide employees with knowledge, skills, competence and expertise in specific subject matters or job functions”. Bigelow [21] also argues that in some cases, companies retrain employees several times without finding out the root cause of employee’s inability to learn or comply with the requirements.

New technological innovation requires employee training. Warren et al [22] indicate that fast and single fire technology has been incorporated into the tile industry since 1990. However, the sanitaryware and tableware sectors delayed in considering such installations. It is in this instance where employee training plays an important and crucial role. Day et al [22] indicate that mass production has dominated this industry for many years, where manual labour, semi-skilled workers and predominantly unskilled labour were the typical scenario. Day et al [22] also stressed that under the new flexible specialisation environment (a lean characteristic), core and periphery workers, multi-skilled artisans and semi-skilled operatives were linked by a chain of sub-contractors. Again, here training plays an extremely important role.

Soft investment is an infrastructural factor, which could be associated with the concept of leanness, see [7]. This is the proposition compared in this research. In so doing this study increases the knowledge about lean production. Moreover, under the methodological point of view and due to the fact that there are no similar studies in relation to the existing correlation between soft investment and the implementation of lean production, this study offers opportunities to develop new perspectives. According to Buhler [23], training the workforce is a key challenge and it is necessary to be reconsidered by managers in order to combat global competition. May [24] argues that people must learn in order to develop knowledge, which will empower the workforce to become independent goal seekers, leanness while developing deeper problem solving skills and critical thinking capabilities, which are key
characteristics of the lean principle of multifunctional teams. Needy [25] indicates that companies claim that people are their greatest assets, however, many of them fail to assess skill standards which need to be developed. In line with this argument, Smith et al [26] stress that the implementation of new management practices in industrialised countries has had a significant impact in employee training.

3. METHODOLOGY

The research methodology followed in this study is based on the formulation and test of two hypotheses. These examine the existing relationship between the two components that conceptualise leanness (managerial commitment towards lean production and real changes made towards the lean direction) and the soft investment, which is assumed to be an infrastructural factor associated with these two components. Hence the following hypotheses were developed:

H1: Firms that have made in tandem investment in the supporting manufacturing infrastructure, SMI – (measured by the degree of commitment – DOC) and real changes towards the lean direction (measured by the degree of leanness – DOL) have also made investment to train their personnel (measured by the variable Investment in People – INVESTP), see [7].

H2: firms that have made in tandem manufacturing structural investments – ISM (measured by the degree of commitment – DOC) and real changes towards the lean direction (measured by DOL) have also trained their production managers (measured by the variable managerial skills – MSKILLS).

The two hypotheses formulated, and presented above, for this investigation are summarized in table 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>DOC</th>
<th>DOL</th>
<th>INVESTP</th>
<th>MSKILLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1 (H1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td>Hypothesis 2 (H2)</td>
<td>+</td>
<td>+</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>Tool</td>
<td>Questionnaire 2</td>
<td>Questionnaire 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondent</td>
<td>Top Management (CEO/President)</td>
<td>Operations and Production Managers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.1. RESEARCH TOOL – SURVEY STUDY

The main research tool used in this study of the tableware industry was a survey questionnaire. As the model and methodology of this study have already been explained in detail, see [7] and [27], it will simply be summarized here. It is prescribed that this survey, in adapted format, can be deployed for data collection in future studies testing the leanness of manufacturing firms. This survey was also supplemented with structured short interviews, external and internal secondary data and plant observation to increase familiarity with the tableware industry and triangulate any results derived for the study for validation purposes. The aim in deploying this survey was to examine the relationship between the main components of the Karlsson and Ahlstrom conceptual framework, see [28] for a reference on this framework, the adoption of lean production principles and managerial commitment to lean production. Because the relevant data were not available in secondary form, primary data collected was necessary. The data generated also enabled the investigation and testing of a number of other different research hypotheses.

For the purpose of this paper, the unit of analysis is the firm. The information was obtained from two levels of the organisation: the production and operations managers, and top management (CEO/president). A different questionnaire was used especially designed for each level. Questionnaire no. 1 was directed at production and operations managers and was used to measure the degree of adoption of the principles that comprise lean production. Questionnaire no. 2 was directed at top managers and was used to measure the degree of commitment of top management towards the adoption of the lean system. Both questionnaires were completed in the presence of the researchers. The two questionnaires measured different variables, consequently, they were analysed independently. A short structured interview was performed on the operations managers with the objective of obtaining more
information about the firms and the managers. Additionally, 14 planned visits were conducted. The objective was to observe the production process in more detail. The population was defined as the firms of the sector with 35 or more employees, included in the list provided by ‘Business Link Staffordshire’ under the title ‘vitreous china table/kitchen products’. A list of 45 companies having these characteristics were acquired from this title sector. The basic assumption of this study was that firms surveyed should have a minimum of 35 employees in order for the new paradigm to be viable. The fundamental reason for this assumption is that a typical process in this industry has six stages: 1. Preparation of the clay, 2. Moulding, 3. Drying, 4. Firing, 5. Decorating, Fixing & Refiring, 6. Packing. If at least 5 workers are assigned to each stage this will add up to a total of 30 employees for manufacturing. Additionally, a minimum of 5 administration staff are needed which adds up to 35.

North Staffordshire County is traditionally known as the heart of the tableware industry in the UK as most of the ‘potteries’ and largest firms are located there. Three experts from CERAM Plc, previously called ‘British Ceramic Research Institute’, reviewed the list previously provided by the ‘Business Link Staffordshire’. They excluded firms that had disappeared or gone bankrupt, sister companies and other sectors that were wrongly included in the list (e.g. refractories and miscellaneous products). They also included some firms that were not in the list but in their opinion were relevant for the study. The new reviewed list comprised 36 firms. These firms were contacted and 33 agreed to participate. All firms completed both questionnaires (with the exception of only one general manager who decided not to participate). The rate of response was higher than 90 per cent, which is exceptionally high compared to similar research conducted in other similar industries, see [29].

4. STUDY RESULTS

4.1. HYPOTHESIS 1 (H1)

To test H1 a correlation analysis was performed. This hypothesis assessed the relationship between the degree of managerial commitment towards lean production (DOC) and degree of adoption of lean systems (DOL) in relation to the investment in people (INVESTP), see table 1. In a previous study, the taxonomy of the firm that comprises this industrial sector was performed by a cluster analysis, see [7].

Since the variables DOC and DOL comprise the concept of leanness, they were both correlated with INVESTP. Table 2 shows the results of this analysis. As can be observed, factors DOL and DOC present a high correlation (p<0.01) indicating that lean production was implemented. Also, the same variables are correlated with INVESTP (p<0.05). However, when executing a one-way ANOVA the means were found not to be significantly higher in lean plants than in traditional plants. Out of this analysis we can infer that if we make soft investments in lean plants, however, not sufficient enough to have a significant difference. Nonetheless, H1 was accepted.

Table 2. Pearson correlation matrix between factors DOL, DOC and response INVP (N=32)

<table>
<thead>
<tr>
<th>Factor</th>
<th>DOL</th>
<th>DOC</th>
<th>INVESTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOL</td>
<td>-</td>
<td>0.520 (**)</td>
<td>0.350 (*)</td>
</tr>
<tr>
<td>DOC</td>
<td>-</td>
<td>0.333 (*)</td>
<td>0.333 (*)</td>
</tr>
<tr>
<td>INVESTP</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

(*) p < 0.05; (**) p < 0.01

4.2. HYPOTHESIS 2 (H2)

To test H2 a correlation analysis was performed. This hypothesis assessed the existing relationship between the degree of managerial commitment towards lean production (DOC) and the degree of adoption of the lean system (DOL) in relation to the training of the operations managers, which was measured by the factor managerial skills (MSKILLS). The variables DOC and DOL comprised the operationalisation of lean production so investigated the correlation between managerial skills and both of them. Table 3 shows the results of this analysis.

As can be observed in Table 3, the factors DOL and DOC presented a high correlation (p<0.01) between them, which indicated that lean production had been implemented. Also, it can be observed that DOL and DOC are highly...
correlated (p<0.01) with MSKILLS. Then, one-way ANOVA was performed followed by Tukey’s HSD procedure. ANOVA showed that there are significantly high effects between \((F_2, 27=15.506000)\) for this variable. Tukey’s HSD procedure has shown that MSKILLS factor in lean firms are significantly higher than those of the same variable in traditional firms, which means that lean firms have managers with better managerial skills than traditional firms. Moreover, lean firms also have better managerial skills than firms in transition. This analysis indicates that managerial skills are a key factor in implementing the lean operations paradigm. In this study, when we talk about managerial skills we included the educational level, his/her managerial experience, training in last 5 years and managerial competencies. The managers’ attitude and interest towards this research project in addition to the level of cooperation towards this study was included in this measurement. It could be observed that in addition of his/her level of experience, training, and skills, managers of this firm had a positive attitude and genuine interest in this investigation in contrast to those managers in traditional firms. This is in line with what Weiss [30] argues regarding that the attitude and interpersonal relations are as, or more, important than the capability and ability. These results provided support for H2, which was accepted.

Table 3. Pearson correlation matrix between factors DOL, DOC and response MSKILLS (N=33)

<table>
<thead>
<tr>
<th>Variable</th>
<th>DOL</th>
<th>DOC</th>
<th>MSKILLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOL</strong></td>
<td>-</td>
<td>.520 (**)</td>
<td>0.548 (**)</td>
</tr>
<tr>
<td><strong>DOC</strong></td>
<td>-</td>
<td>-</td>
<td>0.588 (**)</td>
</tr>
<tr>
<td><strong>MSKILLS</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

(*) p < 0.05; (**) p < 0.01

5. Conclusions

Based upon the results of this investigation, it can be concluded that firms that have made higher soft investment are the lean ones. Therefore, H1 and H2 were accepted and shown that there is a correlation between the concept of leaness and soft investment. These results also are consistent with Buhler’s [23] results, who argues that soft investments tend to be a critical element in successful firms. As it can be seen, training is an extremely important issue when implementing lean manufacturing. Due to insufficient empirical evidence, future research is needed to clarify the existing relationship between soft investment in firms and the implementation of lean production found in this study.

References

Realistic Manufacturing System Integration Training with an Assembly Plant Prototype Based on Transportation and Warehousing Systems

Julio Garrido Campos*

Automation and Systems Engineering Department
University of Vigo
Vigo, 36200, Spain

ABSTRACT

The paper presents the experience of education and training in the field of automated system integration. The experience is based on a "manufacturing plant prototype laboratory" where manufacturing processes are represented as goods transportation and warehousing. The paper will present the prototype made of transportations and warehousing industrial automated systems based on different mechanical principles (overhead warehouses, a stacker cranes for boxes, rolled conveyors, robots, etc.). Each system represents a plant process as, for instance: an overhead warehouse to represent a raw material reception and warehousing with the ability to automatically store and retrieve goods, robots to make automatic transfer between systems, a trolley with stop and deviation devices to represent the assembly shop floor, etc. Each system is automated with different technologies: controllers technologies (several PLC's, PC's, embedded systems,...), different fieldbuses and communications (Profibus, ASI, wireless communications, Ethernet,...), etc. The architecture allows, among training in the horizontal integration of different technologies, the implementation of vertical process integration, from production planning, to machine control movements. The paper shows how this "goods movement's oriented lab" is a very realistic representation of a full assembly industrial plant, although any manufacturing process is actually performed.

1. INTRODUCTION

Systems engineering process is based on a fairly standard approach applied worldwide based on sequence of activities from design to manufacturing, installation, operation, etc. Systems design and development costs are usually well known; however, integration, installation and operation problems are more difficult to predict, as well as their influence in the final cost [1]. System integration involves taking independently developed sub-systems and putting together to make up a complete system. Subsystems faults that are a consequence of invalid assumptions about other systems are often revealed during system integration. During system installation, the system is put into the environment in which it is intended to operate. While this may appear to be a simple process, many problems may arise, which mean that the installation of a complex system can take much more time than expected [2]. Once the system has been installed, it is put into operation and undetected problems may arise at this stage because the system specification may contain errors or omissions.

This situation is particularly true when developing complex automated manufacturing systems, where these there “final” activities (integration, installation and operation booting) are critical in terms of the final system cost. While current industrial manufacturing processes have common problems and requirements because of markets and technologies globalization and because they have to face the heterogeneity and growing complexity of technologies into an always changing environment, “manufacturing systems design” is quite challenging since not only the artifact must be modeled from the perspectives of form and function, but so from the perspective of control [3].

* Corresponding author: Tel.: (++34) 986-812610; Fax: (++34) 986-814014; E-mail: jgarri@uvigo.es
Integration in manufacturing is the systematic approach to organize the manufacturing system as a whole, not only at the field level, but also at the management and corporate level, to produce an integrated and interoperable enterprise system. Enterprise engineering models and tools are needed for a seamless integration of business and manufacturing models, to completely describe the information aspects of an integrated manufacturing system. Integration in Manufacturing through “vertical processes integration” means the ability of plant-level production applications and business systems to share information, exchange services and to cooperate using these information and services [4].

Vertical integration have been faced in different industrial sectors by defining “standard or sector normalized” models and integrated into custom-made “vertical” systems for enterprise resource planning (ERP), SCM, PDM, manufacturing resource planning (MRP) or manufacturing execution system (MES) [5][6].

The ISA-95 standard has become extremely visible in the automation world by defining a model for manufacturing operations, including the reporting and analysis functions that are critical to effective manufacturing. The ISA 95 Enterprise/Control System standard defines interfaces between business functions and manufacturing operations functions. It also defines a new reference model for manufacturing operations systems that is rapidly becoming the new model for MES (Manufacturing Execution System) applications [7]. The ISA-95 international standards for the integration of enterprise and control systems are a useful tool when establishing an operating philosophy. These can be used to determine which information has to be exchanged between systems (for sales, finance and logistics and systems for production, maintenance and quality assurance). However, ISA-95 focuses on MES-ERP integration, but does not go into MES and control systems integration. Controllers may use custom/brand communication mechanisms for MES communication, although there are also standard protocols (as for instance OPC, OLE for Process Control), but oriented to data access and transfer, not to the information content description.

In addition to information education, experience is an important aspect of training [8], since the “aim of training is to develop new skills, knowledge or expertise”. Training in manufacturing systems integration may be focussed on vertical integration though the information, and most of the times it is performed over “isolated processes” [9]. However, to realistically experience installation and operation booting phase problems, a more realistic approach should be employed in the training plan.

The aim of the paper is to present an assembly plant prototype based on transportation and warehousing systems which is used for training in integration and installation of complex manufacturing systems. The paper is organized as follows. Section 2 presents the ISA-95 standard architecture which is the base for the prototype structure. Next, section 3 describes the training system, with an overview of the hardware and control architecture as well as the processes that may be implemented. The paper finishes with some final conclusions and a summary section.

2. MANUFACTURING SYSTEMS INTEGRATION.

The ISA-95 framework can be used to define a high level operating philosophy and to summarise and document how each activity will be implemented. Some activities may be manual processes supported by Standard Operating Procedures (SOPs), while others may be implemented using commercial software for engineering control and automation. The ISA S95 Function diagram describes the activities that must take place in order to produce a product (Figure 1). They represent the business processes as a functional hierarchy model with 5 levels, but distinguishes two domains within a manufacturing company: the Enterprise Domain (Level 4 for business planning & logistics) and the Manufacturing Domain (Level 3 and lower for manufacturing operations and control, and process control: batch, continuous and discrete control) [10].

This standard specifies that each level shall provide the functions listed below and illustrated in

- Level 0 defines the actual physical processes.
- Level 1 defines the activities involved in sensing and manipulating the physical processes.
- Level 2 defines the activities of monitoring and controlling the physical processes.
- Level 3 defines the activities of the work flow to produce the desired end-products. It includes the activities of maintaining records and coordinating the processes.
- Level 4 defines the business-related activities needed to manage a manufacturing organization. Manufacturing-related activities include establishing the basic plant schedule (such as material use, delivery and shipping), determining inventory levels and making sure that materials are delivered on time to the right place for production. Level 3 information is critical to Level 4 activities.
As companies now work to improve their efficiency and productivity, these systems must be modified and/or designed to interoperate. They must coexist under the same IT infrastructure. There is a commonly used IT infrastructure that can be extended to connect the various application levels into a corporate infrastructure. The ISA-95 hierarchy model can be applied to the equipment, people, and information systems in a facility. The model defines levels that can be mapped to systems and to networks.

A typical physical view of ISA-95 systems is shown in Figure 1 (left). Figure 1 (right) represents the levels in a four-layer physical network. The layers should not be confused with the ISA-95 levels, but are instead layers of communication access, designed to maximize network uptime and minimize unnecessary network traffic. A bottom layer is added for automation equipment and for connection of the automation equipment to SCADA (Supervisory Control And Data Acquisition) and HMI (Human Machine Interface) devices. The bottom layer, the automation layer corresponds to connections between ISA-95 Level 2 and Level 1 systems. This is added as a specific layer to the industry standard large system LAN design. This layer is added to specifically control automation network traffic and switch configuration in order to meet the real-time response requirements needed by automation. Level 2-3 network communication is localized within the access layer. This localizes real-time data collection and HMI communication and also provides a measure of robustness and standalone capability. Level 3-4 network communication is localized in the distribution and firewall layer. This provides communication between all ISA-95 Level 3 systems and also provides a single point for firewall protection of the operations network.

While the models just define a way to describe the activities of a manufacturing company, they are often also used to define levels for software applications and networks. ERP, SCM, CRM, and PLM are typical Level 4 applications. MES, LIMS, CMM, batch management and WCS (Warehouse Control Systems) are typical Level 3 applications. PLCs, DCSs, SCADA, and HMI are typical Level 2 systems and applications. There is overlap in the application area, with ERP often supporting some Level 3 functions, and DCS systems often supporting Level 3 and Level 2 functions.

While connectivity is the first requisite to systems to get interoperability, research [11] has shown that the use of structured documents is essential to face information data management and share in complex environments. Among other standards for specific domains, as for instance ISO STEP [12] for CAD/CAM/CAPP domain, XML has shown advantages as a method to embed machine semantic information [13].
3. TRAINING SYSTEM ARCHITECTURE.

With the objective of training vertical systems integration (integration between enterprise domain processes and manufacturing domain processes) and horizontal manufacturing systems integration (different manufacturing systems collaborating in an integrated production process), the “Automation and Systems Engineering Department” of “Vigo University” (Spain) has completed along years a “Manufacturing System Prototype” based on industrial transportation and warehousing systems.

![Figure 2: Industrial Plant Prototype Processes.](image)

The prototype represents a whole assembly plant by reproducing a typical sequence of goods (raw material, intermediate and final product) movements and storages. To get the whole system working, all systems have to be horizontally integrated to move a piece from the input point (raw material reception area) to the exist point (final products expedition) (Figure 2). The prototype has the following systems, which may be viewed in Figure 3.

3.1. “PICK TO LIGHT” SYSTEM (S1).

This system is the raw material reception point and represents an automatic aid for the input of raw materials. Incoming raw material would go straight to S2 (raw material overhead warehouse) or could be first filtered and recognized through this process. The system could work on different ways: sometimes it is used in assembly lines to help operators to pick the correct piece, or even in the final packing phase to help operator through the request orders. In the prototype system, it is being used to help in the raw material entrance process. When a reference barcode is read, the system automatically turns on the corresponding button lights to show to the worker which pieces has to pick. The worker picks them and acknowledges pressing the corresponding button (and the system has to turn off the button light). This way, the system knows which reference is going into next warehouse S2 area.

3.2. AUTOMATIC OVERHEAD STORAGE AND RETRIEVAL SYSTEM (S2/S3).

It allows the automatic storage and retrieval of raw materials by means of a system composed of overhead storage bars, monorail trolleys, elevators, etc. The system has a ground entry point, where raw materials coming from S1 or from outside are hung into mobile hooks (a hook with two wheels able of rolling through monorails). An elevator brings down hooks one by one to the ground level warehouse entry point, and an exit elevator do the same to connect S3 with next S4 system by bringing the stored raw materials from the overhead warehouse area down to the ground level exit point.

The overhead warehouse has two accumulation areas, each one with several accumulation bars. The first area is to allow fast raw material reception, and mobile hooks are stored as fast as possible without taking care of the storage order of the incoming raw material (with different types or references) in the accumulation bars. Second area is for automatically store the raw material (hung from the mobile hooks) in a more efficient order depending on the production plan. Each hook may have just one raw material item, and the warehouse has been designed to allow and handle different raw material references at once. If there is more references than warehouse store bars (where hooks are stored in sequence by gravity), references have to be mixed up. Therefore, to get a specific raw material reference order in the second area or at the exit point to support JIT, automatic movements of hooks have to be performed until the corresponding referenced is in a bar exit, so it could be sent to the warehouse exit.

3.3. “POWER & FREE” TRANSPORT SYSTEM (S4).

This is a transportation system to move goods (raw material, intermediate assembly products, etc) between points. This subsystem would represent the assembly line with a sequence of “working points” where assembly operations are performed, although the prototype only have an input point, an assembly point, and an exit point. The name “Power&Free” comes from the mechanical behaviour of the system, where transportation trolleys or units move through a powered puller (in this case a moving rape with pulling hooks, see Figure 4 left down picture), but the power system frees the transportation units when they are stocked in a queue behind a stopper devuce (for instance, in a working point). The system allows transportation units skip a particular working point (See Figure3, central lay-out, system number 4): for instance, if the transported “reference” has not to perform the manufacturing
operation of a working point then it may be skipped, or because of line balance issues if, for instance, one manufacturing operation need double time than the following one, then two working points are dedicated to the first operation (and transportation units chooses one of them to go in), followed by a working point to perform the second operation (and all transportation units have to go through).

3.4. MANIPULATING ABB ROBOT (S5).

This system is a manipulating robot for transferring pieces from S4 to S6. There is no management information in this process. The robot, when receives an “electric Boolean” order, retrieves the good from a S4 transportation unit and moves it to an exit position where a box from next system is waiting. Therefore, just Boolean interconnection with previous and next systems is implemented.

Figure 3: Industrial Plant Prototype. Centre (left: ground lay-out / right: overhead lay-out). Upper pictures (from left to right): S2 elevator, S1 pick to light system (barcode reader and red and green buttons with light ), S2/S3 overhead warehouse. Lower pictures (from left to right): S4 transportation system (assembly line), S5 robot and S7 belt, S6 automatic stacker.
3.5. AUTOMATIC STACKER CRANES FOR BOXES (S6).

This is automatic storing system, where the product is stored in bookcases which are automatically stored in shelves by a two axis transportation system. The system knows the XY position for each rack, and once there, it is able to put and retrieve a box (product container) through a combination of small lift and down movements and the pneumatic expansion and contraction of two fingers to hold the box. This warehouse has two main processes: one is to store pieces from the robot, the other is to get a particular box (with a particular good) and send it to the next S7 system.

3.6. AUTOMATED ROLLERS SYSTEMS (S7).

This is, together with the robot, the simplest system in the prototype in terms of management. The systems interact with the S6 system (stacker cranes for boxes) to get the container with a final product which is being requested. It moves the container to the exit point, where the good it manually picked, and sends the container back to the first point, to be transferred back to S6. To allow dealing with multiple orders simultaneously, there is an upper line for boxes going from S6 to the picking point, and a lower line for returning boxes.

4. INTEGRATION

From the automation point of view, each system is isolated from the others. Each one has its own control system, field-bus and MMI equipment, as Figure 4 represents. Systems S2 and S6 use Profibus as field-bus technology to connect controllers and field elements, while S1, S3 and S6 use ASI field-bus technology. The use of open field-buses as device network brings a lot of flexibility regarding with controllers. Then S1, S2, S3, S5 and S6 are automated with Beckhoff PLC’s or embedded PCs, but S2 could also be automated with a Siemens S7200 PLC, and S3 and S5 with Siemens S7300 PLC’s or by a PC with a real time operation system (RTX) (Figure 4).

Figure 4: Industrial Pant Prototype: Control Hardware and Network Systems

HMI and SCADA systems may be implemented in PC’s connected with the control system through an “automation network” (Figure 1 right: an Ethernet local network) or through dedicated communications depending on the controller type. All these HMI systems are connected with next level processes: the MES systems to generate production orders and track them (Level 3 of Figure 1 left). The system integration training system may be completed with business processes to production planning as ERP ones. A basic ERP “prototype” application may
be done by simply sequencing production order requests in a Database, coming from different systems: for instance, orders could come from a PDA application using a Wireless network, from outside the system through a web service, etc. Also, other high-level involving all processes (S1-S7) may be implemented, as for instance traceability systems (to record all relevant data about the manufacturing process) or maintenance support systems to handle maintenance interventions (Simchi-Levi and Kaminsky, 2003).

Figure 5: Industrial Plant Prototype Processes organization

To implement these software integrated architecture, working groups of two or there students are responsible of different application. Six groups are responsible of control and MES applications for systems S1 to S6 (Figure 5). Additional groups perform the ERP system, PDM and web services to generate production orders, and the maintenance support system.

Software may be performed in two complexity levels. If the whole system works with a unique “final product reference”, transportation and warehouse systems basically have to move goods from point to point, without having into account any other parameter. S1, S2/S3 and S6 systems would just need to maintain occupancy maps to known which systems parts are free or not. However, a more real integration may be trained if the system works on a “multiple reference” basis, which means that the system has the ability to deal with multiple product references at the same time, each one made of different raw material and with different set of manufacturing operations. In this case, S2/S3 software systems have to maintain a much more complicated record on the goods location (map), and control algorithms would need be more complex, with the objective of being able of storing mixed references and put a specific reference in the output point at a given moment (JIT). Also, S4 would have to decide which products have to perform an assembly operation or not, and therefore, also a “software map” would have to be implemented. Finally, S6 would also need to know which final reference is in each warehouse container/position, etc. Other services as traceability could be also implemented around the “multiple product references” to being able, for instance, of checking if the traced data (raw material used, operation performed) is coherent with the current reference. Integrating all these systems means, not only having to maintain an automatic goods flow from the system input point (at S1) to the output one (at S7), but also to integrate each transportation and warehouse system with business processes (ERP) in order to perform a specific production sequence on time.

Finally, only with realistic system installation and operation booting situations may be trained, and to experience their deep and costly consequences. For instance:

- Delays in the integration and operation booting of one system directly affect the others.

- Failures in the automatic material transfer between systems due mechanic or software error. Software errors may be corrected though “maintenance options” in the applications, but the system has to be properly designed to deal with mechanical failures. For instance, with the ability of detecting and inform about mechanic failures and with the implementation of recovery mechanism.

- Failures in one system “reference software map” may be propagated to the next systems. Coordinated recovery mechanism should be taken into account.
5. SUMMARY

Implementing an industrial automated manufacturing system is a mixture of technologies, control application, management and planning applications, networking and communications systems, data models and formats, etc. Sub-systems have to work together with high performance requirements and integration and operation (booting) phases are key system development phases, as important or even more, in terms of cost influence, than previous designing and implementing phases. However, these phases are difficult to realistically reproduce on academic facilities and laboratories. A intermediate solution as training platform is using real industrial prototypes or machines, where vertical integration may be trained while developing the automation, the Man Machine Interface, and even higher level functionalities and/or processes. The particular case described in the paper, reproducing a whole factory through the goods movements, is a very realistic representation of a full assembly industrial plant, although any manufacturing process is actually performed.

The architecture presented in the paper allows, among obvious training in vertical process integration, from production planning, to machine control movements, allows horizontal integration of different technologies and processes. With the prototype, realistic problems of installation and booting phases may be reproduced to properly understood and training. The prototype sub-systems described in this paper (S1 to S7) have real counterparts on industrial automotive companies in the laboratory surrounding area. Theses sub-systems are not only real industrial systems, but equivalents ones may be physically visited.

ACKNOWLEDGEMENTS

This work is supported by the Spanish Bureau of Science and Technology MCYT (Ministerio de Ciencia Y Tecnología) as project DPI2006-05772.

REFERENCES

**Engineering Management Programs For Advanced Manufacturing Technologies**

Adji Cissé and David Wyrick*
Department of Industrial Engineering, Texas Tech University

**ABSTRACT**

This paper presents current results from an initial investigation of the engineering management programs located in Canada, Mexico, and the United States; this is the first such survey since 1990. Furthermore, the paper highlights the schools which have programs that emphasizes advanced manufacturing systems management. Survey results indicate that approximately one in eight engineering management programs in North American offer a concentration in advanced manufacturing.

1. **INTRODUCTION**

To the general public in North America, engineers are not commonly associated with managerial positions. In fact, 21% of leading American corporations have a chief executive officer whose undergraduate degree is in engineering [1]. With the increasing change in the business world and the fast growing technological environment, companies are asking for well-rounded engineers who are able to practice both engineering and managerial skills. Hence, to meet the needs of industry, the engineering education system has been pressured to innovate its engineering education program to produce a new breed of engineers: engineering managers. As a result, there is impetus for the creation and growth of engineering management programs (EMP's).

Engineering management is defined by Kocaoglu as “the discipline addressed to making and implementing decisions for strategic and operational leadership in current and emerging technologies and their impacts on interrelated systems” [2]. The typical engineering management student will take engineering courses, along with business classes such as accounting, engineering economics, and management.

Engineering management is a fairly new discipline. It developed first in the early 1900’s when the first EMP was created in 1913 at MIT (the program was then called industrial management). Since the mid-20th century, the discipline has experienced a steep growth curve, as indicated in the studies by Dr. Dundar Kocaoglu at Portland State University conducted in 1976, 1984, and 1990 [2]. This is the last known worldwide survey of the engineering management discipline, despite the continued growth of the field.

The aim of this paper is twofold: first, we attempt to update the state of the engineering management academic profile and second highlight the programs notable in their work with flexible automation and intelligent manufacturing. This will be a follow up to the 1990’s survey, however the scope of the study will be limited to the programs within the signatory countries of the North American Free Trade Agreement, namely in the US, Canada and Mexico. The paper will be looking at the growth of EMP’s in these countries, the degree offered, the academic base of the program and some more general characteristics of today’s EMP’s. The need for updating this survey is to provide a current analysis of engineering management education.

First, we will present the literature review which will primarily be a summary of the findings from the 1990 Kocaoglu survey relevant to our research scope, and supplemented with recent developments in defining the body of knowledge of the engineering management discipline [3]. The methodology used in this research will be presented next, including a discussion of similarities to and differences from Kocaoglu’s method. The results of this research will be tabulated and discussed. Preliminary conclusions regarding the state of engineering management education will then be drawn and recommendations for further study will be made.

* Corresponding author. Tel: 1 (806) 742-3543; E-mail: dave.wyrick@ttu.edu
2. LITERATURE REVIEW

The last comprehensive research done on engineering management education was the one conducted in 1990 by Kocoaglu. That research focused on the worldwide growth of the engineering management field. Programs in 20 countries were surveyed, including the US, United Kingdom, Canada, South Africa, Australia, and Hong Kong. The results came as a follow-up to similar previous surveys. The survey indicated a “steep slope, and there is no indication that it has approached the saturation point” as of 1990 [2].

The 1990 survey included the growth of engineering management as an academic discipline, the academic base of engineering management programs, profiles of students, numbers of alumni, faculty makeup, and research areas of the field. Due to the limited scope of the research, this section will only highlight the key points relating to the number of EMP’s, and the curriculum (as compared to the current engineering management body of knowledge discussed by ASEM).

2.1 GROWTH OF ENGINEERING MANAGEMENT PROGRAMS

Kocaoglu’s survey confirmed that as of 1990, there were 121 EMP programs in 20 countries, up from 84 in 1984 and 32 in 1976. This was a growth of 44% in six years, and nearly 400% over 14 years. This dramatic growth had a particular pattern: it was mainly detected in the graduate programs. The survey of the EMP showed that the graduate program had quadrupled since the mid-1970’s. This was in stark contrast to the undergraduate programs, which had increased only 20%, growing from 25 programs to 30. [2]

It is important to note that Kocaoglu included a variety of program titles in his list (17 different types). Sixty-four programs were in engineering management or management engineering. While some programs were rather close, such as engineering administration or MSIE with an engineering management option, others were included that are now considered outside of engineering management. These programs include technology management/management of technology, systems engineering/information technology, and construction management.

Table 1 provides the distribution of the EMP degrees in the US and other countries as of 1990. Of the estimated 11,000 engineering management students, more than 60% were at the MS level, 34% were undergraduates, and the rest were pursuing their PhD’s.

Table 1. Degrees Offered by 121 EMP’s in 1990 (adapted from Kocoaglu, 1990)

<table>
<thead>
<tr>
<th>Degree</th>
<th>USA</th>
<th>Other countries</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.S.</td>
<td>29</td>
<td>9</td>
<td>38</td>
</tr>
<tr>
<td>M.S.</td>
<td>76</td>
<td>23</td>
<td>99</td>
</tr>
<tr>
<td>Ph.D.</td>
<td>19</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>TOTAL</td>
<td>124</td>
<td>42</td>
<td>166</td>
</tr>
</tbody>
</table>

2.2 CURRICULUM OF ENGINEERING MANAGEMENT PROGRAMS

Kocaoglu described the scope of engineering management as having two dimensions: the life cycle dimension and the systems dimension. The undergraduate curriculum at the time typically had two years of fundamental engineering, one year of management, and one year of engineering specialty courses. He described the typical graduate curriculum as having core prerequisite courses (“statistics, simulation, software engineering”), fundamentals delivered in the core and electives (operations research, accounting, financial management, project management, etc.), specialization mainly through electives, and emerging areas.

In 2007, Dr. Donald Merino of the Stevens Institute of Technology discussed the engineering management body of knowledge (EMBoK). EMBoK was developed based on the curricular content of
Flexible Automation and Intelligent Manufacturing, FAIM2009, Middlesbrough, UK

engineering programs which were either accredited by ABET, Inc., or were certified through the American Society for Engineering Management (ASEM) certified EM Programs. He found that curricula fell into six categories: qualitative and conceptual; quantitative and methodical; accounting, financial, or economic; project related; functional management; and engineering and science. Table 2 highlights the major EM curriculum sections and typical courses that are currently recognized as standard EM body of knowledge. [3]

It should be noted that Merino’s recommendations in 2007 represent a refinement of the engineering management curriculum as it has evolved since Kocaoglu’s 1990 study. Many of the course titles are the same in both papers. This consistency indicates that the profession has reached a certain level of maturity in the past two decades.

Table 2: EMBOK, the Engineering Management Body of Knowledge (adapted from [3])

<table>
<thead>
<tr>
<th>Major Functional Definitions</th>
<th>Subfields</th>
<th>Typical Course Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative/Conceptual</td>
<td>Individual, people oriented</td>
<td>Individual psychology, Personnel management</td>
</tr>
<tr>
<td>Courses</td>
<td>Organization or group oriented</td>
<td>Organizational behavior, Management Theory, Teaming</td>
</tr>
<tr>
<td>Quantitative/Methodical</td>
<td>Quantitative</td>
<td>Statistics, Operations Research, Decision Theory, Simulation</td>
</tr>
<tr>
<td>Courses</td>
<td>Methodical</td>
<td>Systems engineering</td>
</tr>
<tr>
<td>Accounting, Financial, and</td>
<td>Accounting, Finance</td>
<td>Managerial accounting, Financial accounting, Cost management, Engineering accounting, Financial management</td>
</tr>
<tr>
<td>Economic Courses</td>
<td>Economics</td>
<td>Engineering economics, Macro economics, Micro economics, Managerial economics</td>
</tr>
<tr>
<td>Project Related Courses</td>
<td>Project Management</td>
<td>Project management</td>
</tr>
<tr>
<td></td>
<td>Capstone</td>
<td>Capstone, Special projects</td>
</tr>
<tr>
<td>Functional Courses</td>
<td>Functional Technical Management</td>
<td>Operations management, Quality management, Engineering management, R&amp;D management, Marketing management</td>
</tr>
<tr>
<td></td>
<td>Functional Business Management</td>
<td>Marketing, Engineering Law, Management Information Systems</td>
</tr>
<tr>
<td>Engineering and Science</td>
<td>Engineering Courses</td>
<td>Anything with “engineering” in the title, except for “engineering management”, “systems engineering”, “industrial engineering”</td>
</tr>
<tr>
<td>Courses</td>
<td>Science Courses</td>
<td>Mathematics, Physics, Chemistry</td>
</tr>
</tbody>
</table>

3. METHODOLOGY

The information in this updated review of engineering management programs was collected through web searches, review of university literature, and the use of Peterson’s Guide [4] and a database of engineering management programs provided through the American Society for Engineering Management (ASEM) [5]. First, a list of EMP programs and the universities offering them was gathered through Peterson’s Guide to Graduate Programs [4]. A second list was made using the ASEM database. Through subsequent web searches, each list was updated as individual programs were either verified or rejected.

This search term used during the internet research was “engineering management.” This meant that some of the titles of the programs included in Kocaoglu’s work were not included in this paper. One notable field that was not considered is construction management. Also, programs in systems engineering and industrial engineering that did not include an engineering management option were excluded. Once each database’s list was confirmed, the two were consolidated into one comprehensive spreadsheet.

Once the list of schools was formed, a survey questionnaire was sent out to those schools, in order to confirm whether they in fact had an engineering management program. In order to gage the role of
advanced manufacturing, the questionnaire asked whether the program was one with an emphasis on manufacturing management. This part of the investigation was primarily to pinpoint the school that offered a manufacturing management emphasis. As this survey is still in progress, the results presented here are preliminary.

4. RESULTS AND DISCUSSION

This section summarizes the results obtained from this research. The interpretation of the results will be made through their comparison to the previous survey results in 1990.

4.1 MANAGEMENT OF ADVANCED MANUFACTURING

Recently, manufacturing has seen tremendous change. Low skill, labor-intensive industries have experienced cutbacks, while new, high-tech manufacturing industries are growing and searching for qualified employees. The skills needed by workers in manufacturing have changed as have the skills needed by its leadership. The conventional educational curricula for engineers must therefore change along with it. Technical skills provided as the core of almost every engineering degree program are still critical, but the workforce needs to come into industry equipped with additional skills. Indeed, a pool of technical professionals with effective management skills and leadership abilities must be provided [6].

In 1997, the Society of Manufacturing Engineering (SME) had already pinpointed the then existing gaps between academia and the industry’s needs. Figure 1 summarizes these as a collection of both manufacturing or engineering skills and managerial or business skills [7]. In general terms, these are similar to the Engineering Management Body of Knowledge [3], although there is the pronounced emphasis on manufacturing.

![Figure 1: Ideal Manufacturing Engineering Skill Set](image)

Though many of these gaps on the engineering and manufacturing have been filled, still today there is a need for manufacturing engineers to develop leadership and managerial skills. Often, this lack of managerial skills in the graduate manufacturing engineers is met through the acquisition of an MBA. Engineering management is, however, another option for acquiring the managerial skills. Indeed, there exist engineering management programs which have an emphasis on the management of manufacturing systems.

As a result, part of the survey was aimed at highlighting universities that have an EMP with a concentration in advanced manufacturing. Of the 36 survey respondents to date, 97% claimed to have an engineering management program. However, the results of the survey shows that a small subset of the programs actually have a manufacturing emphasis; only four of the 36 (i.e. 11%) said they offered a concentration in manufacturing management, as shown in Figure 2.
The four schools that had the concentration in advanced manufacturing were: Oakland University, Eastern Michigan University, Milwaukee School of Engineering and Lamar University. All these programs are hosted in the engineering school, except the one from Milwaukee School of Engineering, which is housed in the Rader School of Business.

Perhaps two of the most prominent manufacturing management programs are Massachusetts Institute of Technology’s Leaders for Manufacturing Program (LMF) and McGill University’s Master in Manufacturing Management (MMM). (These programs have not yet replied to the survey, so this information has been retrieved from their websites.) MIT’s LMF is a two-year program, at the end of which the students earn an MBA as well as an MS in engineering. This program was created to feed the manufacturing industry’s needs for effective managers who had the engineering knowledge as well as the breadth of management skills. The program is comprised of coursework, seminars, workshops and an internship. It is mainly geared toward honing the leadership skills of future manufacturing managers. Figure 3 shows the “academic triangle” used at MIT.

McGill’s MMM program is only 12 months in duration. It is geared towards professionals with an engineering or science background who wish to become effective leaders in the manufacturing world. The program is the result of the partnership between the engineering and management school. The program is structured similarly to the LMF program with coursework, seminar, workshop and a semester-long project in industry during which the student apply their expertise to real-world challenges. One main makeup of this program is its intense focus on managerial skills [9].
The survey also asked the number of manufacturing courses required to fulfill the program, and it was found that an average of 5 courses was mandatory (Table 3). This was greater than the required manufacturing courses for the generic engineering management program which had on average has 2 courses.

Table 3. Mandatory Manufacturing Courses for Concentrations within Engineering Management

<table>
<thead>
<tr>
<th>University</th>
<th>Program Name</th>
<th>Manufacturing Courses Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oakland University</td>
<td>MS Engineering Management</td>
<td>3</td>
</tr>
<tr>
<td>Eastern Michigan University</td>
<td>Engineering Management</td>
<td>6</td>
</tr>
<tr>
<td>Milwaukee School of Engineering</td>
<td>MS Engineering Management</td>
<td>5</td>
</tr>
<tr>
<td>Lamar University</td>
<td>MS Engineering Management</td>
<td>3</td>
</tr>
</tbody>
</table>

4.2. Recent Growth of Engineering Management Programs

The results from this research are summarized in Table 4. As of late 2008, there are 159 EMP’s reported in 2008 in the USA, one in Mexico, and eight in Canada. Of those, we were able to verify 153 programs in the US, none in Mexico, and 8 in Canada. These results, compared with those from 1990, show growth of about 63% in the US and 33% in Canada. This is a considerable growth in both countries, as depicted in Figure 5. As for the reported program in Mexico, it could not be verified; further studies must be conducted to determine the status of that program.

Table 4: Distribution of EMP’s By Country, 2008

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>159</td>
<td>153</td>
<td>94</td>
<td>63%</td>
</tr>
<tr>
<td>Canada</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>33%</td>
</tr>
<tr>
<td>Mexico</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Total NAFTA</td>
<td>168</td>
<td>161</td>
<td>100</td>
<td>61%</td>
</tr>
</tbody>
</table>

Similar to the 1990 survey, the present survey highlights a prominence of masters’ degrees over the bachelor’s and PhD’s offered. As shown in Figure 5, in the US as well as in Canada, the masters’ degree is offered far more often than the bachelor degrees: 121 masters offered over only 21 bachelors in the US, and six masters’ degrees over one bachelor offered in Canada.

Table 5 shows the change in the number of programs at the bachelors, masters, and doctorate level in the United States since 1990. Data were not reported at this level for Canada and Mexico from the 1990 survey. On its face, it would appear that the number of masters programs in engineering management has increased by almost two-thirds. This trend of increasing masters’ degree offerings was noted in the early 1990’s survey as well [2]. Table 5 indicates that there are now fewer programs at both the undergraduate and doctorate levels; however, it is not clear how many programs were included in the 1990 survey that were not included in this survey (e.g., construction management).

Table 5: Change in Number of Programs at Bachelor, Masters, and Doctorate Level

<table>
<thead>
<tr>
<th>Level</th>
<th>USA</th>
<th>Canada</th>
<th>Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MS</td>
<td>125</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>PhD</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 5. EMP Degrees Offered by Country
Table 5. Engineering Management Programs in the United States

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B.S.</td>
<td>21</td>
<td>29</td>
<td>(-28%)</td>
</tr>
<tr>
<td>M.S.</td>
<td>125</td>
<td>76</td>
<td>64%</td>
</tr>
<tr>
<td>Ph.D.</td>
<td>15</td>
<td>19</td>
<td>(-21%)</td>
</tr>
<tr>
<td>Total</td>
<td>161</td>
<td>124</td>
<td>30%</td>
</tr>
</tbody>
</table>

4.3 ADDITIONAL FINDINGS

In the course of conducting the survey of engineering management programs, we identified several changes and differences in the programs in contrast to the findings of 1990. One major improvement is the way in which the courses are delivered. With the advancement of technology, namely the breakthrough of the worldwide web since the late 1990’s, education has taken a new form; more and more courses are done through distance education. Engineering management has been a leading engineering discipline to join this trend; as such, many EMP’s are currently offered online.

It is important also to note one major distinction. In this survey, construction management was not included as an EMP, whereas it was in Kocaoglu’s survey; four out of the 6 non-verified EMP’s in our survey are program in construction management. The reasoning behind the exclusion of construction management is, though it has a number of engineering management aspects, it is a more specific type of engineering program, one restricted to civil engineering (one exception is the University of Wisconsin-Madison, which has an ABET-accredited program in engineering management). This may be an explanation for some of the decline of the bachelors’ degrees from 1990 to 2008. Perhaps, a stricter or more comprehensive definition of engineering management must be established for this type of survey in order to increase consistency of results.

Five engineering programs have become accredited by ABET, Inc., since 1990, bringing the total number to ten. These are listed in Table 7. [10] Also since 1990, the American Society for Engineering Management has begun a certification process for masters programs. This certification process includes a self-study document and site visit. Table 8 gives the ASEM-certified masters programs and the year of initial certification. [5]

Table 7: ABET-Accredited Engineering Management Programs [10]

<table>
<thead>
<tr>
<th>Institution</th>
<th>Program Name</th>
<th>Degree</th>
<th>Initial Accreditation</th>
<th>Next General Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force Institute of Technology</td>
<td>Engineering Management</td>
<td>MS</td>
<td>2004</td>
<td>2009-10</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>Engineering Management</td>
<td>BSEMgt</td>
<td>2005</td>
<td>2010-11</td>
</tr>
<tr>
<td>Missouri University of Science and Technology</td>
<td>Engineering Management</td>
<td>BS</td>
<td>1979</td>
<td>2008-09</td>
</tr>
<tr>
<td>North Dakota State University</td>
<td>Industrial Engineering and Management</td>
<td>BS</td>
<td>1971</td>
<td>2012-13</td>
</tr>
<tr>
<td>Oklahoma State University</td>
<td>Industrial Engineering and Management</td>
<td>BS</td>
<td>1936</td>
<td>2009-10</td>
</tr>
<tr>
<td>University of the Pacific</td>
<td>Engineering Management</td>
<td>BS</td>
<td>2004</td>
<td>2012-13</td>
</tr>
<tr>
<td>Rensselaer Polytechnic University</td>
<td>Industrial and Management Engineering</td>
<td>BS</td>
<td>1978</td>
<td>2007-08</td>
</tr>
<tr>
<td>Stevens Institute of Technology</td>
<td>Engineering Management</td>
<td>BE</td>
<td>1992</td>
<td>2009-10</td>
</tr>
<tr>
<td>United States Military Academy</td>
<td>Engineering Management</td>
<td>BS</td>
<td>1985</td>
<td>2008-09</td>
</tr>
<tr>
<td>University of Wisconsin-Madison</td>
<td>Construction Engineering and Management Option in Civil Engineering</td>
<td>BS</td>
<td>1995</td>
<td>2012-13</td>
</tr>
</tbody>
</table>

As for the constitution of the faculty, the majority has a background in engineering. There was no noticeable trend as far as full-time versus part-time faculty; depending on the institutions, there can be either more part-time faculty or more full-time. Results from the initial review of Peterson’s Guide, the ASEM database, and internet search for engineering management programs were presented at the 2008 ASEM Annual Meeting [11].
Table 8. ASEM-Certified MSEM Programs [5]

<table>
<thead>
<tr>
<th>Institution</th>
<th>Program Name</th>
<th>Initial Certification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Dominion University</td>
<td>Masters of Engineering in Engineering Management</td>
<td>2003</td>
</tr>
<tr>
<td>Stevens Institute of Technology</td>
<td>Masters of Engineering in Engineering Management</td>
<td>2003</td>
</tr>
<tr>
<td>Stevens Institute of Technology</td>
<td>Executive Masters in Technology Management</td>
<td>2003</td>
</tr>
<tr>
<td>University of Missouri-Rolla</td>
<td>Master of Engineering Management Program</td>
<td>2003</td>
</tr>
<tr>
<td>St. Cloud State University</td>
<td>Master of Engineering Management</td>
<td>2007</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

Engineering management as a discipline is still experiencing significant growth, particularly in the United States. In the US alone, the total number of EMP’s in 2008 has far exceeded the number found in the worldwide survey of the programs in 1990: 153 programs in the US now versus 121 worldwide in 1990.

The graduate programs, especially the master’s degrees, mainly drive the high growth of EMP’s. The reported number of bachelors programs, on the other hand, is declining in numbers. Having said that, the number of ABET-accredited programs has doubled since 1990.

The structure of engineering management programs seems to be reaching a standard. Most of the programs found in both the US and Canada closely followed the structure depicted in the EMBoK.

This research relied heavily on publicly available data. Further work is justified to supplement the picture presented here. Future work should make use of direct measures through direct interaction with each program. Such effort should provide a much better understanding of the makeup of the profession, including questions of faculty, students, alumni, and research. It is strongly recommended that the term “engineering management” be in keeping with ASEM and ABET definitions.

Overall, the field is steadily growing, as evidenced by ABET-accreditation, ASEM-certification, and the Epsilon Mu Eta honor society.

REFERENCES

An intelligent design pedagogy for Engineers.

Keelin Leahy and William Gaughran *
Department of Manufacturing and Operations Engineering.
University of Limerick,
Limerick, Ireland.

ABSTRACT

Preferential Learning Styles (PLS) are very seldom considered in engineering pedagogy, and often the approach to design strategies is limited to experiential discovery learning. However, when the pedagogical approach and learning style are not compatible, learning is not optimised. The project explored the effects of implementing a web-strategy approach to design activities of engineering and technology students, with particular emphasis on the active and visual preferential learning styles and ‘right brain’ cognition. A web-based strategy focused on developing aspects of students’ design ability and creativity through design activities, which acknowledged the students’ PLS. The design strategy was evaluated by a comparative analysis of control and experimental projects completed between, September 2007 to May 2008. The subjects for this study consisted of one hundred and seventeen undergraduate students. The control project followed a ‘traditional’ approach, and the experimental project was underpinned with a web-strategy. A comparative analysis of the two groups revealed significantly improved design skills, when the subjects PLS influenced the strategy. This paper discusses the methodologies applied and identifies key outcomes to further refine and develop a ‘best practice’ web-strategy for design activities pedagogy.

1. INTRODUCTION

Studies on students’ preferential learning styles (PLS) have discovered that the PLS of students within the Technologies education programmes are similar to those in third level Engineering programmes[1]. However PLS may alter or further develop depending on the learning environment one is exposed to. Thus a varied pedagogical approach to design and creativity must be implemented to facilitate students developing his/her full potential and capabilities.

The design tasks within the web-strategy aspire to developmentally appeal to and increase the participants’ design, creative, and problem-solving aptitudes. In current practice in design activities, the language and practice of design is complex and unclear. It was anticipated that the web-strategy would provide the foundations necessary to develop transferable skills which improved the learners design ability, creativity, as well as the understanding of the application of the principles, practices and processes for successful design activities within Irish third level educational institutions.

2. PREFERENTIAL LEARNING STYLES.

For this study the Felder-Silverman model was used to determine student’s preferential learning styles (PLS). This model, known as the Index of Learning styles (ILS), is a 44-item forced answer instrument, initially developed by Richard Felder and Barbara Soloman and finally proven validity in 2005 by Felder and Silverman [1]. The characteristics for the various domains of learning styles are detailed in table one.

* Corresponding author: Tel.: (353)61202850; Fax: (353)61202913; E-mail: Keelin.Leahy@ul.ie
Table 1: PLS characteristics.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Retain and understand information best by actively doing something with it.</td>
</tr>
<tr>
<td>Reflective</td>
<td>Think quietly about the new material.</td>
</tr>
<tr>
<td>Sensing</td>
<td>Like learning facts. Solve problems using well-established methods. Practical.</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Prefer discovering possibilities and relationships. Innovative. Dislike repetition.</td>
</tr>
<tr>
<td>Visual</td>
<td>Remember best in what they see illustrations. (Majority are visual learners.)</td>
</tr>
<tr>
<td>Verbal</td>
<td>Get more out of words.</td>
</tr>
<tr>
<td>Sequential</td>
<td>Gain understanding in linear steps, each step following logically from the previous one.</td>
</tr>
<tr>
<td>Global</td>
<td>Learn in large jumps: absorb material randomly without seeing connections until suddenly.</td>
</tr>
</tbody>
</table>

Customising the pedagogical approach to suit the students learning needs is important. However as the student group has a range of learning styles it is also important that the pedagogical approach is not totally focused on the dominant learning styles. Therefore the strategic pedagogical approach to design education and training will appeal to many learning styles, while the dominant learning styles will be the focus. This will result in reforming the ‘lecture’ approach pedagogy in third level education creating an improved design ability of students.

3. CURRENT DESIGN PEDAGOGY IN THIRD LEVEL INSTITUTIONS.

Traditionally Ireland is not seen as a design literate nation. It is predominantly known as a musical and literary profound nation. However, nowadays design can be made applicable to such a vast range of areas that its meaning too has become vast and difficult to simplify. However for the context of this research design refers to design in teacher training in Third Level Education Institutions in Ireland. We are all designers. Whether it is what we will do for the day to what we want our living room to look like. Design could be seen as making changes. According to Mark A. Runco “teachers are responsible for many of the experiences which can dramatically influence the creative expression.”[2]

In 1976, Ken Baynes commented that the central concept regarding design creates a lot of confusion for many educators in design education. [3] In the context of teacher training this confusion is evident in aspects of the present Irish second level education. This issue is also highlighted by Ai Girl Tan, 2007, that “not all teachers including those who are expected to develop creative potentials of learners posses a comprehensive view of creativity” or design. [2] Another issue concerning design education is the background or profession the knowledge or experience is derived from. This can lead to different views and practices regarding the meaning of design. According to Lawson, “design can be seen as an attempt simply to solve a local problem or to improve a particular situation.” [4] Friedman defined design as a process. [5] Lawson et al reinforce this idea by stating “the design world would argue that design is itself also a process of discovery, of learning and even a form of research.” [4]

Design has many different domains, which is the case with the design element in the second level (high school) technologies. Similarly education is an extremely broad discipline. Design educators need to be professional in their design practice and educational theory to convey effective design pedagogy to students.

The ‘design process’ used in most second level educational institutions is linear. The correct practice for a design process is outlined by Fiell; “the design process is not linear; it is rather a complex activity similar to a game’s strategy but strangely it is a game where the rules are continuously changing and that is what makes it so fascinating and mysterious.” [6] The linear approach contradicts the reason behind why one ‘designs’. A design process evolved primarily as a problem solving process due to a problem. Design on the other hand does not occur due to a problem but rather due to ill-defined problems. [7] However design processes implemented in a problem solving methodology in education systems are causing difficulty as they are not “accurate descriptions of actual practice.” [4] A strategic approach attempts to aid trainee teachers influence and improve students design ability and creativity through awareness to preferential learning styles.
4. STRATEGIC APPROACH TO DESIGN EDUCATION.

The technologies are unique subjects within the Irish second level curriculum. They involve cognition, reflection, development of ideas, and testing of ideas in a practical mode [7]. The development of a strategic approach to design in second level education is required to cater for one of the many disciplines within the technologies.

From discovering the current pedagogical approaches to design education and the PLS of students within the Technologies a strategic approach to education and training has been devised to cater for the needs of students in second level education. The strategic approach consists of a web-strategy resource, which provides the framework for vital material content and approaches deemed necessary to design and the development of creativity. The strategy primarily focuses, though not wholly, on the preferred learning styles of students in the Technologies. This approach attempts to overcome the issue of design being implemented in a predominantly ‘lecture’ or verbal setting as is the norm practice in many third level institutions. This ‘lecture’ or verbal approach is a key impediment for right mode creative activities [8]. The purpose of the web-strategy is to provide trainee teachers will a pedagogical approach and training to design education with a ‘grounding’ or permanence within second level education. It also aims to develop background information to design thus increasing the user’s design ability and creativity. This is achieved by the various aspects the web strategy encompasses, including history of designed artefact, concept development, thinking like a designer (processes, practices and principles) and sketching as a design tool (graphic ideation, communication, modelling etc).

5. KEY ASPECTS OF WEB-STRATEGY.

Design encompasses all facets of life. However acknowledging and being aware of the vast knowledge of design relevant to design education is initially essential to ensure a clear understanding of the practice of design is established. Establishing the key aspects of design education occurred to ensure the relevant aspects were in fact influential toward design understanding and ability development. The key aspects deal with history of designed artefact, concept development, thinking as a designer and development of sketching techniques, which involve lessons and tasks for the promotion of design capabilities. The lessons provide introductory reinforcement of the key concepts of design. The design tasks aspire to developmentally appeal and increase the participants’ creative aptitude by emphasising on the active and visual learning styles and the R-mode. Table 2 explains the key aspects in more detail.

<table>
<thead>
<tr>
<th>Skill set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept development</td>
<td>Deals with developing the R-mode toward idea development. This is achieved by initial sketches developing participants’ confidence and motivation in their sketching abilities.</td>
</tr>
<tr>
<td>History of designed artefact</td>
<td>Highlights the importance of acknowledging the development of products through history and the significance of design through history.</td>
</tr>
<tr>
<td>Thinking as a designer</td>
<td>Explores the methodologies and ‘secrets’ to design in the form of principles, practices and processes of design.</td>
</tr>
<tr>
<td>Development of sketching techniques</td>
<td>amalgamates many of the key aspects of the web-strategy. Sketching is a vital tool for designing. It allows one to develop and communicate their ideas. Sketching is a key aid to the exploration of concept development and thinking as a designer.</td>
</tr>
</tbody>
</table>

These key aspects of the web-strategy were applied and key outcomes were identified which will lead to further refinement and development of a ‘best practice’ web-strategy for design activities pedagogy. The application of skills derived from the key aspects will be the greatest obstacle. Many students have become programmed in terms of ‘what to think’ and are vast bodies of knowledge. However, students are not being challenged sufficiently in terms of the application of knowledge or ‘how to think’ in overcoming situations or problems.
6. INTELLIGENCE POTENTIAL.

The Irish third level education teacher training is a restricted system in Ireland with only one established training college. Another training college has been established however this programme is still to be accredited a degree programme by the Department of Education and Science as it is only in its third year of establishment. The Irish third level education system is largely dominated by ‘the points system’. For a student to obtain a course of their choice they must achieve the minimum points acceptable to enter the course of preference. Achieving the entry points to a course though does not guarantee entry onto a third level course as demand may be greater than the course capacity. Thus pressure is peaked for most students during the leaving certificate cycle to achieve.

Pressure is peaked as students must consume as much information as possible to achieve their chosen course of study in third level. The leaving certificate examination mainly requires students to demonstrate knowledge and comprehension of a subject area. However the application of knowledge to solve a problem or issue is not assessed at present in the leaving certificate examination. Students are absorbing vast quantities of information in a ‘jug and mug’ system which is not beneficial for the practice, comprehension or application of design. For the subject of design, the approach of the experienced practitioner leading the notice, a similar practice of professional designers and their apprentice’s decades ago. However this tutor-apprentice practice for design education does not exist in the Irish third level or second level education mainly due to the lack of awareness by educators in terms of the professional practice of design. The pedagogy in existence in third level institutions is primarily based on what to teach rather than how to teach. This is having a knock on effect in the second level education system. At present design pedagogy is based on “what to think” rather than “how to think” thus design ability and creativity is not being nurtured to its greatest capacity. [9] The design web-strategy sets out to build intelligence potential amongst educators and students on the application of knowledge toward the practice of design and increase the skills necessary to improve design ability and creativity.

7. METHODOLOGY.

PARTICIPANTS.

The methodology for the strategy implementation focused on third year students in a third level teacher training programme for the technologies. The technologies include Material Technology Wood, Technology, Technical graphics and Engineering. The participants for this study consisted of one hundred and eighteen undergraduate students. The participants have a mean age of 20.8 and standard deviation of 2.55. The breakdown of male and female students is a ratio of approximately 23 male: 1 female.

DESIGN and PROCEDURE.

The trainee teachers were surveyed initially to determine their preferential learning style. The first design project occurred by trainee teachers applying their own ‘traditional’ approach, this was called the control project (P1). P1 was used to determine students approach to design and also their current level of design ability and creativity. P1 occurred in semester one third year of the trainee teachers’ degree programme. The experimental design project (P2) followed and was underpinned by the web-strategy which was influenced by PLS of students in the technologies. P2 occurred in semester two third year of the trainee teachers degree programme. P2 involved initial guidance and instruction from the web-strategy. The web strategy lessons and tasks were carried in two hour laboratory sessions per week over an eight week period. A factor of experience was determined by comparative analysis of the two design projects. The factor of experience revealed significantly improved creativity and problem-solving skills.

Control Approach – Project one (P1).

All participants in the study initially completed design project one independently. Prior to initiating the project trainee teachers were given a pre-project survey to determine their interests, aptitudes and current level of ability. The stage of completion for this study only required participants to design to final design selection stage. The trainee teachers could create a 3-D prototype if so wished. Realisation of the final design did not occur to reduce the focus
An intelligent design pedagogy for Engineers.

on the manufacture of the design project. The trainee teachers were given a thematic design brief which they independently, without peer instruction, completed the design or redesign. On completion of the design project the trainee teachers were then required to complete a post project survey. This served the purpose of determining students’ level of understanding, interest, evaluation and assessment of their design project.

**Experimental Approach-Project two (P2).**

On completion of the control approach design project-one participants took part in a strategic approach to designing for design project-two. The experimental approach was peer lead and controlled to ensure the strategic approach, which comprised of various developmental stages, was standardised throughout trainee teacher groups. The stage of completion is the same as per design project-one, final design selection and prototyping. Analysis of the strategic approach is carried out by multiple choice questionnaires on the completion of each design task to determine the level of contribution to the overall development of design ability and creativity of the trainee teachers. The strategic approach was prescribed via instruction and support from the web-strategy. The various stages of the strategic approach comprise of concept development, thinking as a designer, history of designed artefact and development of sketching techniques.

**8. IMPLEMENTATION OUTCOME.**

**PREFERRED STYLE**

For the initial preferential learning style testing a range of second level students (525), of the technologies, were surveyed. From this study, the findings correlated with the findings of Felder [1] and Seery [10]. The PLS of second level students in the technologies are the *Active* and *Visual* learning styles, as illustrated in Figure 1. On comparison third level trainee teachers were surveyed to determine their preferential learning style. It can be observed that the PLS correlate for students and trainee teachers of the technologies.

![Figure 1: Preferential learning style comparison; second level and third level.](image)

However the sequential preference (53%) / global (47%) suggests a style not suited to design activity. Sequential thought is linear. Global thought is the ability to view the larger picture, a skill required in design activity. The sequential preference could be due to the ‘lecture’ style approach to learning in third level institutions with the modular system of learning. This sequential preference is alarming and also highlights the preference, though erroneous, for implementing a linear approach to design activity in second level design education.
EFFECTS OF A STRATEGIC APPROACH TO DESIGN EDUCATION.

Overall the web-strategy received a positive appraisal and outcome from the third level trainee teachers. The main aim of the web-strategy is to increase teachers understanding of design and ability to implement and practice design successfully in the second level education system. On a broad level there was a greater understanding of design activities achieved by seventy-five percent of participants. (Table 3: T1) The highest influential key aspect to design ability is ‘development of sketching techniques’ as illustrated in Figure 2.

Table 3: Effects of strategic approach.

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Indifferent</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>No comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T1</strong> Clearer understanding of practices of a designer</td>
<td>13%</td>
<td>62%</td>
<td>13%</td>
<td>4%</td>
<td>-</td>
<td>8%</td>
</tr>
<tr>
<td><strong>T2</strong> Observed Great Improvement in Sketching Skills from Start to Final Sketch</td>
<td>29%</td>
<td>33%</td>
<td>23%</td>
<td>5%</td>
<td>2%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>T3</strong> Web strategy developed your design ability and creativity</td>
<td>10%</td>
<td>36%</td>
<td>38%</td>
<td>11%</td>
<td>1%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Figure 2: Benefit of design key aspects

To validate the influence of the development of sketching techniques from the web-strategy there is an overall increase in trainee teachers sketching ability. Trainee teachers noted an improvement in their sketching skills with approximately sixty-five percent noting an improvement in their sketching ability from the strategy implementation (Table 3: T2). This improvement is evident in the sample work from trainee teacher’s initial and final sketch during the web strategy intervention in Figure 3.

The effects of the web-strategy skills are evident in the application to P2 (Table 3: T3). Overall trainee teachers noted improvement and ease to convey their design ideas. This improvement is graphically visible in trainee teachers sample work in Figure 3. Initially for P1 many students copied or traced the object they were aware of, or created symbolic shapes such as cars or trucks. Many students noted difficulty during P1 to think of original ideas or communicate these ideas. However during P2 trainee teachers demonstrated more unusual ideas and more confidence when designing.

The positive influence of the web strategy in terms of promoting trainee teachers’ design and creative ability was noted by approximately eighty-five percent of the cohort (Table 3: T3).
An intelligent design pedagogy for Engineers.

To reinforce the positive influence of the web-strategy on the majority of trainee teachers, there was a vast improvement in the end of semester grading between P1 and P2, as illustrated in Figure 4. The factor of experience between semester one (P1) and semester two (P2) is clearly evident, with a positive improvement in design ability and creativity for semester two (P2).

![Initial concept sketch](image1.png) ![Final concept prototype](image2.png)

Figure 3: Sample work from trainee teacher’s initial sketches to final sketches or prototype during the web strategy.

![End of semester 1 and 2 project grades](image3.png)

Figure 4: End of semester 1 and 2 project grades
9. DISCUSSION.

One main obstacle noted by second level students is “how can I design when I have never been taught how” (fourth year, second level student). Similarly from an educator’s perspective much of the learning associated with design cannot be verbalised. However design is not a verbal subject. From observing and interviewing professional designers it is predominantly a visual and active activity. Another issue hindering the practical activity of design, according to Uichol Kim, is that “in modern classrooms, teaching methods have become increasingly routine and objective in the transfer of knowledge.”[2] This explains why students design intellect is being hindering as they are not being taught ‘how’. Knowledge acquisition for assessment is the driving force in the Irish education system. Design education cannot survive under such an environment. Design is not solely about the ‘transfer of knowledge’. Design deals with the application of knowledge to solve an issue or problem. Applying knowledge in design education can be achieved through visual and active means.

10. CONCLUSION.

Creating a suitable environment to promote learning is becoming a greater difficulty in today’s society due to many reasons, one example is technological advancements. However as Vygotsky noted “How do we educate the child raised in a world of instant information, where interactive technologies have led them to believe they can act on the world with the press of a button?”[11]; educate the ‘how’ rather than the ‘what. Teaching the subject matter of design is deemed by many people as ‘impossible’ as many believe it is an innate ability. However design can be nurtured. A web based design strategy, which caters for the needs of a range of students with different preferential learning styles, can have a significant impact on their design and creative abilities. Recognising the ability to gain or improve design ability was exhibited during this study. Overall there was an increase in trainee teachers design ability and creativity. Education is a constantly evolving area, improvements and advancements are constantly required to provide for the needs of students. The web-strategy will undergo continual refinements to improve students and educators understanding of design activities in the technologies. The methodologies applied during the web strategy intervention identified the key skill set necessary for intelligent design pedagogy for design education and training in engineering and technology educational environments.

11. REFERENCES


MANUFACTURING TRAINING SYSTEMS IN DEVELOPING COUNTRIES - A LIBYAN CASE STUDY

A. AFSHOUK AND P. W. SHELTON

1 Head of Faculty, Postgraduate College Janzzur, Tripoli, Libya.
2 Assistant Dean Education Partnerships, SST, University of Teesside Middlesbrough, UK, TS9 6HX

ABSTRACT

Attempts have been made to explain and describe the effective elements of training and development systems, and its contribution in organisational development cycles, and performance to achievement of organisational goals. Research literature has identified the link between training and development systems and employees’ overall satisfaction, top management support of training, and manager’s involvement in the training function. Most of these studies have been conducted in developed countries, which are completely different (culturally, economically, and politically), from the developing countries e.g. Libya.

Survey data was obtained from the employees (subordinates and leaders) working within the largest non-oil company in Libya. A significant relationship was found between overall job satisfaction and elements of training and development systems. Selection methods for training courses, assessment methods of training needs, time spent in training, training method, kind of training, training places, and training provider were determined to be significant in their relationships to training satisfaction.

The study concluded that there was a positive relationship between the elements of the training and development system function and the surrounding environmental factors, such as job description, job satisfaction, motivational systems, promotional systems, and supervision style. Respondents did not show high satisfaction with the chance to obtain new knowledge, skills from the job, or practice of their skills on the job. Respondents to this survey did not place high value on the training and development within their organisation. Leadership viewed the constructs more favorably than did employees. Based on these findings, recommendations are presented for practitioners in the fields of training and human resource development, as well as for managers of employees working in public companies.

INTRODUCTION

The concept of work-based training has been widely examined to try to explain and contribute to our understanding of the role of such training to develop employees’ attitudes and behaviours as a factor in job satisfaction. The literature on work attitudes in particular has identified and examined the relationship between training, job satisfaction and productivity improvement e.g. by preparing manpower for change in the work environment. Figure 1. Represents how the training management function operates, and who whilst having input to the design and delivery, concentrates on the first and last steps of training.

Identify Needs Design Deliver Evaluate

Figure 1. Stages in the training Process.

In the case of work based training, the ultimate goal of any training procedure is to transfer the knowledge, skills, abilities, or behavior acquired in a training event into improved performance. Inferentially, the failure of
any training activity to achieve this aim may be attributed to any of these stages, not least being inadequate initial identification and analysis of the organisation’s needs.

Gagne and Medsker [1] highlight this objective: improvement in workplace performance is a consequence of the transfer of learning from classroom to the job setting. That organizations can effectively help trainees transfer learned skills to new situations is of growing importance. Ford & Kraiger [2] among others, agree on the importance of needs analysis, though highlighting the role of both implementation as well as evaluation. The success of training is dependent upon correctly following and performing all steps of the process including previous analysis of training needs development, implementation of an adequate training plan and final evaluation, (Antonio, et al, [3]).

Whilst there would be few organisations that would deny the value of training and the amount of training and development appears to be increasing, surveys and expert opinion suggest that spending for training and development is not evenly spread across organisations and employee positions. Black and Lynch [4] examined survey data and found that eighty-nine percent of U.S. employees had received no training, small organisations with fewer than 100 employees were much less likely to provide formal training than large employers (1000+ employees), organisations using high performance work practices such as Total Quality Management and Benchmarking were more likely to offer training to their employees and better educated workers, managerial and professional employees, are more likely to receive employer based training than other employees. Brinkerhoff and Gill [5] furthermore contend that most organisations never assess whether the money spent on employee training and development has actually advanced the organisation’s overall productivity or improved individual performance.

Rothwell [6] adds weight to the argument that failure to assess the effectiveness of training implies that value for money is not being determined or by implication obtained. As a result of the concern for this lack of accountability of training and development systems, literature such as Carnevale [7] Robinson and Robinson [8], Zemsky and Oedel, [9], McClelland [10], Brinkerhoff [11] and Brinkerhoff & Gill [5] had focused on the mean responsible elements behind ineffective training and development systems, and investigated what was going wrong with training and development systems in private and public organisations.

Critics assign the main reasons behind the failure of training and development systems to the influence of some key factors such as:-

- Senior managers lack of support for training and development systems even where they support the attainment of the organisation’s missions Carnevale [7], Robinson and Robinson [8], Broad and Newstrom [12], Brinkerhoff and Gill [5].
- Senior managers view the training functions as unessential, Zimsky and Oedel [9].
- Training budgets are usually the first to be cut at times of economic difficulty.
- The absence of an understanding in many organisations of the connection between training and achieving the organisation’s goals, Brinkerhoff and Gill [5].
- Most organisations determine training needs reactively rather than proactively, usually around perceived immediate job-based shortages or short-term forecasted knowledge and skill needs,
- Where training success is considered, it is often by the numbers enrolled on the course, not whether training achieved the organisation’, or target employee’ needs, Brinkerhoff, [11].

A further cause of failure can be attributed to a combination of change resisting human nature and failure of immediate recalcitrant supervisors or resistant managers or even antagonistic co-workers to support or reinforce the newly acquired skills, attitudes or behaviours, Hawthorne, [13].

Broad and Newstrom [12] contend that “...most training investments do not produce full and sustained transfer of new knowledge and skills to the job”. Brinkerhoff [11] cites studies that show that only 8 -12 percent of what trainees received in training programs transfer to the workplace, as translated into improved attitude, behaviour, and job performance.

In a literature review of training evaluation practices for the period 1969-1986, Foxon (1989) concluded that there was a widespread under-evaluation of training programs and that the quality was not consistent. The need for measurement of training effectiveness is often referred to, but there is few good examples of rigorous evaluation of training programs”, Foxon, [14]. He found that 75% of organisations surveyed do not go further in evaluation efforts, because they do not know what else to do. Moreover such measures of training “success” fade after trainees have returned to the job, and other measures of success, such as the use of the new skills on the job or increased productivity, are ignored, [13].
At the same time results of the empirical studies on Libyan management in general, and training and development in particular, point out that the training in public organisations is run and applied with no effective systems, and missing acceptable standards Agnaia [15], Almehdi [16]. Despite the huge amount expended in training, Libyan state owned companies have difficulty in identifying their training needs, which has led to a common feeling to overcome the existing shortage of skilled people in public organisations, and dissatisfaction among employees that they do not have enough required skills in the job Agnaia, [15].

Furthermore, the decisions related to training activity are still dependent on many cultural factors, which are far away from scientific methods such as personal relations, family ties, tribalism, etc. rather than on an established procedure, [15], [16], focus on the assessment of training needs and selection for training programs in Libyan state oil companies. They explored how these companies select their employees for training programs and the main ways and techniques used by the companies to perform this activity.

Finally, organisations rarely assess the health of their entire training and development system, to determine how to sustain and continuously improve it. Many practitioners would concur with Bishop [17] that insufficient effort is devoted to studies conducted at the organisational level which examine how training fits into the organisation’s overall competitive strategy and affects its profitability.

**Methodology**

This study covers one public company, The Libyan Iron and Steel Company (LISCO). The choice of this company was made because of the following points;

- It the largest state-owned industrial company in the country.
- The number of employees is larger than any other company, about “7000” employee giving opportunity to survey a significant number of respondents
- The company’s employees represent different educational levels, some of them educated out of country.
- The chairman of the company is an academic being a staff member at Al-Fateh University and sympathetic to the aims of the study.
- The company have a large training centre with capacity of 600 trainees.
- The company is state-owned but will transfer to the private sector under the government’s privatisation plan.

The following techniques were selected to be the study’s instruments of data collection:-

- The self-completion questionnaire has been selected as main instrument.
- Some personal observations as supplementary data.
- Interviews with certain employees and leaders.
- Documentary evidence.

Two types of self-completion questionnaires were designed; one for employees and the other for leaders. The design philosophy of the two types of questionnaires was based on the fact that they had to be simple, clear, and understandable for the informants whilst on the other hand they had to be integrated, efficient, and easy to analyse.

Both work group questionnaires were divided into three main sections, each section included some subsections. Both questionnaires contained sections I, and II (Demographic and Organisational characteristics respectively), being very similar. The major differences occurred in the last section (Section III - Training Aspects and Support), in terms of the number of questions and the intent.

In terms of credibility of the survey, it was felt to be important to get the different viewpoints of the two different roles. Clearly the leadership have a very different function, having key influence as in any organisation (by virtue of their position), not least of which is the impact of their interpretation of company policy and even personal understanding of and commitment to the organisation's training strategy.

The responses received from Section III, were used to further explore the extent of the elements of best practice described in the Company Training and Development Framework: Best Practices Framework (Human Technology 1993) [18] and elsewhere in the literature:-

- How the training system was integrated into the company strategic goals.
• Existence of explicit administrative policies for training and development
• Formal procedures for the assessment of training needs.
• Formal procedures for assuring the transfer of training to the job

The measurement method that has been used in this study was connected to an operational definitions concept, following Nachmias & Nachmias, [19] using operational definitions as measurement procedures to bridge the conceptual-theoretical level with the empirical level.

Although the 500 responding employees represents only 7.1% of the workforce, it was considered a significantly large sample, whereas although only 50 managers responses were obtained, this represented 38% of this category. Some 30% of employee responses were not returned which is related to employee suspicion and lack of trust of the motive for this type of exercise. On the other hand, the received number of the management questionnaires was only 12 % short of the number issued. This was influenced by the effect of unavoidable developed personal relationships with the author and the influential effect that the recommendation letter exerted on managers carrying, in comparison, less weight with subordinates.

A number of standard statistical techniques were used as appropriate to analyse data for this study, depending on the types of variables concerned. Descriptive statistics such as measures of central tendency (mean, median, mode) measure of dispersion, and frequency distributions were used to describe the important aspects of the study variables. Reliability and correlation analysis were used to measure the association among the variables. Cross-tabulation was used to examine the association between variables, and chi-square ($\chi^2$) computed to assess the statistical significance of the association in the cross-tabulation. The Pearson-product moment correlation coefficients test was also computed between independent and dependent variables. All the responses were processed and analysed through the Statistical Packages for the Social Sciences 12.0.

**EVALUATION**

The extent of the survey precludes presentation of data in the raw form. These may be found elsewhere Afshouk [20]. Each section included scale and subscale questions in the specific areas the questionnaires were designed to investigate, personal and organisational characteristics, employees' satisfaction with the job, the work environment and training and its relationship to the company perception of the achievement of the elements of effective training.

Some personal data was collected following Olugboyega [21] in research about training in Nigeria who emphasised the impact on determination of training needs. Several scale items were obtained on each of the following, personal characteristic, job and positional details and significantly job satisfaction.

The main purpose of any training program is to cause change in the attitude and behaviour of the target people, overcoming incompetencies or developing some new knowledge or skills. This is an important area to obtain feedback on acquired behaviour or skills. The literature indicates that there are existing some indicators of poor behaviour in Libyan public oil companies [16]. Reading across from this experience prompted examination of the following factors in Section II, Job Characteristic (16 scale items), Work Environment; (14 scale items) and Existence of Poor Work Behaviour (10 scale items).

The work examined a wide range of factors, but perhaps the most interesting was in Section III where the differences in attitude where examined more closely. Data was collected from employees about the training activities and systems in the company including Previous Training (4 scale items), Training Method (3 scale items), Training Evaluation (4 scale items), Selection Method (6 scale items) and Training Transfer including 10 scale items

This cluster seeks to explore the change in the employee's attitude and behaviour as a result of training. Similar factors were evaluated from the management point of view except an evaluation of management involvement:-

**III-K. Manager Involvement in Training; includes 6 scale items.**

Groups of scale items were considered in each category to determine correlations and for given scale factors e.g. sixteen from the Training Satisfaction cluster broken down into 12 scales items from Sections I, and II such as:-

**ESCE1C7, LSCE1C8.** The chance to get obtain knowledge and improve your skills.

**ESCE2A5, LSCE2A5.** Opportunities for acquiring job-related skills.

**ESCE2A6, LSCE2A7.** Extent to which you developed skills and knowledge.
ESCE2A7, LSCE2A8, Opportunity to learn and acquire experience from your job.
ESCE2A15, LSCE2A15, The existence of a variety skills in your job.
ESCE1B6 LSCE1B6, Are you satisfied with your acquired skills.

And:-
4 scale items considered in section III:-
LSCE3A1, Do you consider your self qualified enough as manager.
ESCE3E5, Are you satisfied with your skills obtained from training.
ESCE3C2, LSCE3C1, How do you evaluate the course.

These assess respondents' satisfaction with the training they have received and the extent of utilisation in the work place. Responses were acquired from employees and leaderships’ opinions. Further clusters were evaluated as follows

Attitude toward training
Training's role in achieve the organisation's task.
The Top-management Support of Training programs.
Determining training needs.
Manager's involvement training activities.
Evaluation of training.
The training transfer in job place and effect on the employees' behaviour.

BASIC TRAINING PRECEPTS

From both sides of the management divide, from both employees and supervision and management there was much positive response to the concept of training. It was clear from responses that there was a clear correlation between the number of courses received and educational level and unsurprisingly age, gender, and job position. Those with University education were most likely to be on the receiving end of further training, with most having been on at least two courses. This is consistent with the western context where it has been shown that better educated workers, managerial and professional employees are more likely to receive training than other employees [4].

The majority of training was on-the-job (70.7%), while 34% were trained by internal staff, 98% were trained inside the company, consistent with comments from the interviews indicating that consideration of the training cost was strongly taken into account when the company planned training courses. On the other hand, leaders training courses were designed short-term courses less than a week 58.8 and mostly off-the-job (86.8%). 76.3% were provided by external staff, and in contrast to subordinates, 65.8% trained outside the company.

Employees, however were found to prefer long term course training preferably provided by external staff, from outside the company, and off the job training. Manager’s perceptions were to the exact opposite however.

71 out of 150 employees have been asked to evaluate the course, but only 6.8% of by company training management, the remainder by the courses provider. The management however were more likely to be asked for evaluation by the company, some 25%. The training provider may be more inclined to evaluate the course, but it does beg the question as to the impartiality. The training provider is likely to seek or “engineer” a positive response to his efforts?

The findings indicate that all employees were suggested for training by their managers on the basis of 81.3% have been selected by their supervisors’ recommendation, 7.3 by their record, and 11.3% thought because it was their “turn” to train. 76.3% of leadership were suggested for training by their boss and 23.7% by more senior management. Conversely 15.8% state that they selected their subordinates on the basis of their records, and a majority of 58% state they select subordinates according to their personal opinion. It may be concluded that managers involvement in the selection for training procedure, may lead to a recommendation to top management to target them for more specific training courses about training management.
TRANSFERENCE TO THE WORKPLACE

The ultimate goal of any training activity is transference of learned skills or abilities to the job. Training effectiveness is gauged by the degree to which training results in changes in job-related behavior. This complex task could be achieved through the manager’s observation of the acquired change in their subordinate’s behavior. A range of scale items in leadership’s questionnaire were utilised to explore whether skills and knowledge learned in training activities are transferred to work, whether leadership were held accountable for ensuring change of their subordinates behaviours, attitude toward work, and skills, e.g.:

- Clear understanding of job duties
- Clear understanding of the company policies
- Attitude of satisfaction with hard work
- More punctual
- Greater belief in the Company goals,
- More trust in my supervisor
- More cooperation with colleagues
- More ability and initiative applied to the job
- More care about job quality and the company reputation, (commitment)
- More confidence and enthusiasm for work.

Frequency analysis was employed to describe the items cluster distribution. The findings indicate that training positively affects employees and improves their understanding of duties (60.5%), understanding the company’s policies (55.3%), punctuality (52.6%), more trust in supervisor (78.9%), more cooperate with others (65.8%), and greater concern for the company’s reputation (71.1%). These results confirm that Libyan employees are engendering a positive attitude toward training as concept, which is transferred into the workplace. One may conclude Libyan managers could employ this attitude to overcome the public companies’ inefficiency, and more readily achieve the target goals.

The leadership expressed satisfaction with the chance to get new knowledge. Furthermore, the respondents thought that the availability of learning opportunity was “…gave the opportunity for acquiring job-related skills” (55.3%), “great opportunity to increase the extent of skills development and knowledge from the job” (57.9%), “an opportunity to learn and acquire experience” (52.6%), “very great extent of job contribution to personal growth” (57.9%), and “great extent of variety of skills job requires” (57.6%).

The employees’ standpoint was characterised by satisfaction with the chance to get new knowledge, while they saw a very great opportunity for job contribution to personal growth and development, and a great chance for acquiring job-related skills”, and developing skills and knowledge from the job.

On a more negative side, the analysis of the scale items in terms of the “management involvement in training” cluster gave indicators that the leadership were not involved in the training activities at all nor were they involved in determining the subordinates training needs. Similarly they were not involved in delivering any courses, did not share in designing any courses, and were not involved in evaluating the training. Their opinions were not considered by the top management or the training management, even in response to a training budget. It may be interpreted that the managers do not recognise how their subordinates prefer to train. The managers should be fully involved in the training and development activities, which will help to build more effective training and development systems. Buckley and Caple [23] highlight, “…to build an effective training and development system, it is contended that managers must be involved in all the training stages and adopt a wider view of training as an integral part of the whole organisational system”.

The lack of top management support for the training and development systems in Libyan industrial organisations was found to be similar to what has been found in the Western organisational setting. Here critics argue that one of the key ingredients of a successful training and development system is often missing, [5], [7], [8]. It is apparently clear within the responses of the items, such as the missed link between promotion and advancement to high employment grade or position, non-consideration of the employees or managers opinions, non-involvement of managers in the training function, and the absence of a reward system linked to training. The interviews with such employees and managers in the company support this results, where they indicate that the top management deal with training as an unnecessary activity, a necessary evil. On the other hand, in strategically oriented training, content is selected or designed to provide employees with the knowledge and
skills they will need in the future, Rothwell and Kazanas, [24]. In this study, the most commonly addressed training area were technical, quality, and job skills as the most frequent content of training.

**MOTIVATIONAL ISSUES**

Further negativity was experienced from employee’s responses to the degree of motivation they received to train, strongly in agreement with an equal supervisor lack of motivation to train them. 72% did not feel that their supervisor motivated them to train and 96.7% thought that the promotion system do not link-up with their skills or performance. 69.7% doubted that training or acquiring new skills would offer them chance to advancement to high position. Even if they did well in training, there was no linkage to advancement. Training was not even necessary for them to remain in their job. In comparison 74% the leadership were satisfied that they adequately motivated their subordinates, 76% believed that the promotion system did depend on skills and training and that these were very important for employees to remain in their job.

These findings indicate that there is clear disagreement between subordinates and leadership about the training motivation. Subordinates felt poorly about the training motivation they received. Their belief in the role of training may signify that employees train to satisfy their personal desires more than work benefits e.g. “promotion, reward, or remaining in the job”. Enhancement of these features could lead to motivation of subordinates to practice what they have learned in the job, as they see the purpose of it, Ashton, [25]. The latter indicated that the main factor of supporting workers learning was “…The necessity for support of the learning process in terms of feedback to the learner, and of the need to reward the individual for the acquisition of skills if the process is to continue through time”.

Comparisons between subordinates and leadership opinions signify a failure of managers to understand the direct link between training and motivation. It is related to their involvement level in training needs determining, and any method of training needs analysis in the company, leading to a gap between training activities and the company's needs.

The fact that respondents in this study believe that the training system enjoys a relatively low status in the organisation echoes the opinions of both training and development directors in the organisation from training experts who argue that in the majority of organisations, the training and development system does not enjoy high status. The latter has been a major area of concern for training and development practitioners for a number of years [5], [8]. It would seem reasonable to suggest that the most critical factor for a successful training and development system is that it must be elevated to a position in the organisation from which it can make a significant contribution to helping the organisation achieve its strategic goals. The areas that are so often the subject of discussion and research in the field of training needs assessment, training evaluation, content, methods are the tools and processes for helping the training perform a strategic organisational role.

**CONCLUSION**

As organisations undergo change, critics argue training systems must also change to help employees meet the challenges of a new work environment. Organisations, the arguments go, must elevate the status of the training and development system in order to make it a tool for organisational change. It is no longer enough for the training and development system to be merely the source of a menu of courses that bear little if any relevance to what the organisation needs employees to know and be able to do it must instead be an instrument of change. To be an instrument of change, top managers must be involved in the process, for “without their involvement, it is impossible to make the changes that are necessary for highly effective training” [5]. A training system that meets today’s needs will not necessarily meet tomorrow’s needs. The training system of any organisation must be modified and continuously improved.

Equally importantly, organisations can no longer afford to provide training that has not been evaluated for its contribution to their own strategic goals and mission and its effectiveness and use on the job to achieve those goals. This study confirms what practitioners and experts have long recognized as a weakness in training and development: the effort to make sure that training knowledge and skills transfer to the job [5], [12].

Respondents’ negative perceptions of the effectiveness of training and development offered by the company is perhaps the most serious of the findings for the organisations Effectiveness goes to the heart of what training is all about in an organisation: giving employees the knowledge and skills they need to perform their jobs effectively, [24].
An effective training system has an impact on employees’ behaviour on the job. Many factors contribute to whether what is learned in training activities leads to improved work performance: employee accountability for applying new learning and manager accountability for making the workplace conducive to using new skills and knowledge. Employee perceptions that mechanisms are not in place to help them use what they have learned puts the worth of training in question.

REFERENCES


Enhancing the Training of Engineering Learners

Sajid Abdullah*

Institute for Automotive & Manufacturing Advanced Practice (AMAP)
Department of Computing, Engineering and Technology
University of Sunderland
Sunderland, Tyne & Wear, SR5 3XB, U.K

ABSTRACT

North East U.K businesses involved in design are seeking to recruit graduates with good CAD knowledge and engineering competencies. Higher education establishments within the United Kingdom (UK) have traditionally focused on an instructor led approach; where a lecturer instructs a group of learners, this approach is typical within engineering faculties, delivering modules in the subject based around computer aided design (CAD). The important factor in any training is to offer knowledge in a motivating way. Higher education establishments within the UK are also faced with issues of improving student retention and enhancing the student experience. Higher Education is tackling and coming to terms with these issues, and one such step in that direction is the adoption of e-learning approaches. In response to the needs of industry, and faced with the demands of teaching quality and enhancing the student experience. It is important to implement best practices into higher education teaching and learning methods, recent advances in the use of screen-capture based instructional technology have stimulated the investigator’s research interest towards exploring the use of such technologies in an attempt to improve teaching delivery and enhance the student experience. This paper summarises the research into development of curricula, pedagogy and teaching methodology of a CAD module. The major aim of this study explores the possible use of the e-learning material “Solid Professor” based on screen capture software in the delivery of CAD training to engineering learners and the effectiveness of current teaching methods in ordinary CAD training.

1. INTRODUCTION

Academics and practitioners are addressing the theories of learning and instruction with the aim of understanding learning. Increased understanding of learning should, in turn, enhance the design of instruction and lead to more effective teaching practice.

Curriculum developers are influenced by a diverse number of factors large and small, which makes bringing together of these influences difficult to achieve. A curriculum has to take into account, for example, the demands of industry, government and other interest groups at a major level, and such factors as the interests of learners, the abilities and preferences of instructors, the organisational structure and available resources at minor level.

The two greatest concerns of employers today are finding good workers and training them. The effective performance of an expert CAD user, which is earned over years of experience, is the outcome of continuous learning and professional development.

One of the factors that curriculum developers and instructors consider when creating or implementing training is establishing whom they will teach. Learners change: they are different now to 5 years ago, and in turn to 10 years ago, 15 years ago… Today’s student is seen as techno-savvy, time-pressured, used to ‘creativity’, ‘collaboration’, ‘networking’…, concerned about engagement, interaction, connection with the faculty, potentially going into employment that expects them to hit the ground running. Higher education establishments within the UK are also faced with issues of improving student retention and enhancing the student experience.

* Corresponding author: Tel.: (0044) 191-5153392; Fax: (0044) 191-5153377; E-mail: sajid.abdullah@sunderland.ac.uk
CAD software is changing and being upgraded so frequently that it is difficult for practitioners, let alone educators, to remain current and proficient in their use. This leads to a situation where instructors are often learning the CAD software just prior to providing in-class demonstrations.

Furthermore, each learner has their own pace of learning making it difficult to teach CAD related technologies in a stand-and-deliver format, where some students excel and subsequently are waiting on the instructor, and others lag behind and may never grasp the intended learning objectives. Instructors need an easy and quick to-use instructional technique for teaching software applications [1].

With the software applications becoming more complex and visually oriented and learners more distant, teachers need a tool that will assist them in delivering learning objectives and help them with the demands of software demonstrations.

This research explores technologies and teaching practice in delivery of the case study module delivered by the investigator and presents the research pilot study methodology of teaching using the e-learning component Solid Professor [2], which is a commercially available e-learning system. Solid Professor [2] is a screen-capture based e-learning system, for delivering 3D CAD training for solidworks software application.

2.0 OBJECTIVES OF THE RESEARCH

E-learning is becoming a readily available technology in place today, recent advances in the use of such screen-capture based instructional technology have stimulated the investigator's research interest towards exploring the use of such technologies in an attempt to improve teaching delivery and enhance the student experience.

The major aim of the study explores the possible use of e-learning material based on screen capture software in the delivery of CAD training to engineering learners and the effectiveness of current teaching methods in ordinary CAD training.

Objectives:

- To determine if e-learning material based on instructional screen capture software training material can replace traditional Face to Face teaching in the higher education?
- To determine if e-learning material “screen capture software training” be used to support learning in the higher education context?
- To determine if the E-learning material “screen capture software training” enhances the student experience—improve how our students feel about what we give them.

3.0 METHODOLOGY

At the University of Sunderland, within the Department of Computing, Engineering and Technology (DCET), the B.Eng programmes are modular courses, operating in a full-time mode over a three-year period. Each year consists of two blocks, which are periods of nominally 15 weeks each and there is a mid-block break of one week. Almost all modules in the programme run throughout the year. Modules completed in the first block are assessed and examined within that block and the results published at the beginning of the next block.

The delivery of the Solid Modelling module is through a series of lectures followed by tutorials, 2 hours in length, nominally over 15 weeks. The Solid Modelling module, is a 20 credit module for undergraduate level 2 students in the Department of Computing, Engineering and Technology (DCET) at the University of Sunderland (U.K), which is delivered on the BEng Mechanical Engineering with Design and BEng Engineering Design and Manufacturing Programmes.

3.1 CLASSROOM METHODS “TRADITIONAL”

The Solid Modelling module was delivered to a small group of students, this platform gave the investigator an opportunity to achieve three main characteristics: active participation, face-to-face contact and purposeful activity [3].
The current “traditional” delivery of this module was a student centred learning approach; lectures introduced the design methods and approaches which were later reinforced by the design exercises. The lecturing part informed the theories and rational of the topic being facilitated in the class. Whereas the tutorial sessions focused on introducing the learners to 3D CAD application and problem solving, involving active learning. Tutorials were intended to provide hands on experience to the students working by themselves, with an instructor present.

For the success of these methods, the investigator has conventionally relied heavily on hands-on CAD training approach with ICT presentations—“demos”, PowerPoint’s, and manuals in a corporate training style and problem-based learning. To remain “pedagogically-competent” the investigator communicated the course and class objectives to students.

The assessment developed design skills by the application of the ‘design process’ to a real engineering design problem. The design assignment also provided an opportunity to further develop design knowledge and abilities acquired in design and practical skills and to apply new abilities in solid modelling and environmental assessment techniques.

Student support is a key aspect of higher education (HE), especially in the institutions like University of Sunderland where access/widening participation is a focus on the agenda, for quality output and student retention for non-traditional background students. The considerations include flexible modes of study as well as a supportive environment with focussed support and guidance from tutors and peer support from students. Lecturers and the module leader, are available to contact through e-mails, telephones and drop-in.

3.2 E-LEARNING EVALUATION PILOT CLASS

3.2.1 CONTEXT OF THE PILOT

The investigator was inspired to experiment his teaching methods from a conventional way of teaching to a more modern-student centred teaching style. The main reason being that for many years of his business career, he has focussed on training industrial clients on the best possible ways to use the CAD products and develop best practises. The question arises, was he still behaving like a corporate trainer? The question arose from self reflection.

It is clear from a review of the literature, in particular Honey and Mumford [4]. Who emphasise the way people learn is paramount importance when designing curricula, different styles require different methods of teaching and learning.

E-learning is explored to accommodate distance learning, self-study, and tutoring. Student’s learn at different rates; traditional lectures may be too fast for some and too slow for others. In conventional lectures a student needs to concentrate and remain focused, understand the content at the pace determined by the teacher. The learning curve for different students will vary, some will prefer “less and often” approach, whilst other may prefer longer sessions less frequently. An E-learning environment caters for these individual approaches by giving the student control over his/her learning. The learner can decide when and for how long he wishes to work and seek assistance when he/she requires it.

The study takes a broad view of e-learning, viewing it as any learning that takes place using such screen-capture video based instructional technology. In considering the short time scale of the research and available time, a trial version of the commercially available “Off the shelf” 3D CAD e-learning product SolidProfessor [2], was downloaded and installed for the pilot study, however there are other applications that could have been employed to produce the same output ‘end result’. These may also allow the for greater flexibility in producing learning materials, directly from PowerPoint presentations and other ICT applications to be video captured and constructed to create a learning course system. These applications are Adobe’s Captivate, Camtasia, ‘Snag It’ and another widely known commercial 3D CAD e-learning system ‘I Get It’.

3.2.2 PILOT EVALUATION METHOD

The evaluation of the learning of CAD engineers on the module was performed using both formative, observational and summative methods. At this stage the formative methodology was employed with a group of volunteer learners to “test run” various aspects of the instructional e-learning product Solid Professor [2].

A total of four tutorial sessions were held with a total of twenty one learners. The learner’s consisted of learners from the Solid Modelling module and Industrial learner’s (novice CAD users) enrolled on industrial CAD courses.
The first stage required an cognitive walkthrough of the 3D CAD e-learning product Solid Professor [2], thus the user interface was evaluated in terms of screen layout and user navigation, before the pilot class delivery. It gave the investigator an early evaluation of the e-learning product Solid Professor [2] without involving learner’s “end user”.

This second stage heuristic evaluation identifies usability problems based on established human factors principles. The method will provide recommendations for curricula design improvements that could assist in the use of future course material delivery. [5]

The e-learning product Solid Professor [2] was chosen and installed in an ICT laboratory within the University of Sunderland, for the delivery of the 3D Modelling class. The e-learning material was chosen because it consists of tutorials that provide step-by-step instructions, including both sound and video instruction on how to complete each SolidWorks 3D CAD exercise and at the same time covers 3D CAD principles and content that is directly related to the course learning objectives. Each lesson has specific learning objectives that are listed at the beginning of each chapter providing focus for that lab.

The pilot test learner’s were presented with a brief explanation of how to access the e-learning product ‘Solid Professor’[2], and how to work to complete the on-line demonstrations. After this initial instruction students were directed to follow each E-lesson step by step and to complete the solidworks core concepts learning module. After giving this instruction the instructor was NOT available to assist in the learning delivery and the pilot study relied on the learner’s to follow the screen-capture movies and to learn from this activity at individual pace. During this period the learner’s were observed as objectively as possible. However due to the ethical issues the Instructor was present during the structured activity.

The learners were then asked to complete the questionnaire. The investigator then engaged with the learner’s in a small group discussion lasting approximately ten minutes.

3.3 RESEARCH TOOLS

A questionnaire survey method was used to elicit learner feedback and the views of academic teachers regarding e-learning, during this research study in the last two terms of the academic year (2008). The research questions required the collection of both numerical and narrative. The process of putting the questionnaire into action followed four steps – define purpose, decide information requirements, analyse into component parts and define questions [6]. This is a method of summative evaluation and was useful for determining the attitudes of the learners.

A self completion questionnaire was employed with both open and closed questions with the majority being closed format using Likert scalar format. The purpose of the questionnaire is clear – it aims to obtain information about learning styles deep or surface and the effectiveness of e-learning as a mode of delivery for CAD teaching.

The questionnaire was emailed or hand delivered along with an information sheet introducing the investigator, and in layman’s language explaining the purpose and procedures of the research.

Students at different stages in their studies may have different views about their experiences and expectations of the course or module taught. This cannot be accounted for at present in this research.

An analysis of student and academics opinions based solely on questionnaires lacks depth, and does not allow for an exploration of issues identified by respondents. Interactionist research requires a more active approach, in which respondents are able to explain their views. There are a number of research tools which can be adopted with interviews being particularly popular.[7]

Two academics from another North East U.K university were identified for interview from the survey, the important selection criteria being that the academic’s were existing users of such e-learning material to which this study relates. The interview protocol followed Spradley’s [8] ethnographic interview models.

4.0 RESULTS & DISCUSSION

The pilots classes appeared to have ran efficiently, with no obvious usability issues. The general impressions made by the students during the informal group discussions at the end of the pilot class studies indicated a positive attitude towards the class session and delivery of the material. No major issues relating to course content or the instructional quality of the course were drawn to the investigators attention at that stage. Most however
made informal negative comments such as if they had not been well briefed regarding the general access of the e-learning application i.e. installation, and access to the e-lessons for the exercises, some learners felt this may have discouraged them a little and were daunted by this process.

4.1 Survey Result Analysis

The paper provides a selective analysis of the research survey results. In respect to qualitative data, a number of 15 teaching respondents, and 21 learner respondents to the questionnaire, made known their opinions and views of e-learning and the manner in which this method is delivered within the universities, in particular within the North East of U.K. According to Cohen and Manion [7] a minimum sample size of thirty is acceptable if some form of statistical analysis is to be carried out on the data. As the data sample is low, results need to be treated with caution.

5.0 Summary: Answering the Research Questions

The study is clearly supported by literature, the results of students evaluations can also appear to be contradictory and difficult to interpret [9].

The investigators own belief from this research is that the introduction of e-learning material will not, wholly on its own encourage deep learning. Students are affected by a diverse range of influences to learning issues if the circumstances encourage it. The balance between deep and surface learning may change as the individual learner develops through the stages of the course. Learners appear to develop a more appropriate learning style as the course or module progresses, and at the same time develop a broader conception of the subject matter, suggesting that the current course delivery is achieving desired goals.

5.1 Can e-Learning Material Based on Instructional Screen Capture Software Training Material Replace Traditional Face to Face Teaching in the Higher Education?

The investigator’s opinion is that the e-learning material alone cannot replace traditional face to face teaching, the findings from this research focus towards a blended learning approach.

This is best summarised and supported by responses to the questions, presented in Figure 1, and Figure 2. This is also supported by the responses to the survey question, presented in Figure 3.

![Q53 I prefer face to face teaching](image)

Figure 1. I prefer face to face teaching, using weighted responses.
The views of the academics in the responses made to the open ended questions in the survey and responses made during interviews heavily support the view.

The investigator believes a teacher should be tally involved with the class, dedicated to students and be prepared to devote time and energy for them. The investigator considers the enthusiasm of a motivated teacher rubs off on the students, who take inspiration and encouragement which develops their desire to learn. In addition, a good teacher needs to personalise the needs and problems of the students. This has been observed in the case of a few of the weaker, shy and in some international students who need additional help and but hesitate to ask for it.
5.2 **Can the E-learning Material “Screen Capture Software Training” be Used to Support Learning in the Higher Education Context?**

The pilot learners give the impression from the results presented in Figure 4, and Table 1.0 (appendix) that the use of the e-learning component is overall an effective tool, from the responses in can be observed that the e-learning system has positive results.

![Overall Effectiveness of E-learning Pilot](image-url)

Figure 4. Overall effectiveness of E-learning tool SolidProfessor using weighted responses

The investigator is of the opinion that e-learning is a promising teaching and learning tool to introduce students to engineering and technology. E-learning technology provides another way of understanding engineering CAD methods and experiences, both of which are necessary if students are to construct meaning from their learning.

The following drawbacks were concluded from the survey responses and interview questioning, despite learners finding the e-learning ‘solid professor’ [2] environment easy to learn and use. It was reported by the academics during interviews that if we were to create our own material using ‘camtasia’, captivate, or ‘snag it’, users maybe be recording screen capture video clips within minutes. Despite how simple these applications may be to use, they are time consuming when it comes to development of screen-capture e-learning material. As this requires additional scripting, planning and editing. As indicated in the findings of this research during interviews it took longer to develop screen-capture tutorials compared to the traditional forms.

- “I can’t say how long it takes but it takes maybe three times, four times and up to ten times as long, depending on the construct you want to convey to the students.”
- “I would say five to six hours to produce one hour worth of electronic learning material.”

Those choosing to employ such e-learning materials should allow for this additional development time. This additional development time may be offset by improvements in student learning at later stages of learning process, and time saving to the academic in preparation later on, but no evidence was determined to prove this during this research, and requires further investigation.

5.3 **Does the E-learning Material “Screen Capture Software Training” Enhance the Student Experience—Improve How Our Students Feel About What We Give Them.**

The investigator is of the general opinion that the e-learning material, does enhance the student experience, but will require an effective support framework if implemented. The pilot learners give the impression from the results presented in Figure 4, and the positive responses to question 37, 38, 40, 55 & 56, presented in Table 1.0 (appendix) and consider the use of the e-learning component is overall an effective tool.
Overall, the investigator believes e-learning is an extremely useful teaching and learning aid. E-learning encourages real life situations promoting the development of critical thinking skills. Technology is very useful and should be utilised effectively and with care.

The e-learning material, screen capture video demonstrations assist in the teaching delivery and learning process by providing students with a reason why they are going to perform the task. It transfers the aims and objectives of the exercise (hopefully keeping students interested in the subject, rather than being distracted away from the subject by thoughts of “where is this going”), on the other hand just watching a screen captured video demonstration tutorial can be a bit boring especially with out sound, and may appear lengthy.

The investigator considers that “seeing is believing” and so audio-visual aids like using power point slides for lectures, providing hand outs, showing video clips to emphasise important points, are effective learning aids with proven impacts. In addition, the research findings have formed an impression on the investigator to encourage the use of e-learning applications (with tutor support) to stay in tune with the needs and demands of the learners and industry.

Further research is required to study an instructor led approach that combines both textbook and ‘SolidProfessor’[2] screen-capture based e-learning (blended learning approach) and compare that to the tutorial approach using a learning with an instructional manual.

The research presented has provided me some insights that what can be done to design and deliver students learning such that no one leaves the university without the desired skills and professional development alongside with the successful completion of the degree.

ACKNOWLEDGEMENTS

The questionnaire employed in this research study was a revised version of the Experience of Teaching and Learning questionnaire (ETLQ) constructed by Entwistle, McCune and Hounsell [10] & the characteristics for the learners was adapted from the research of Laidlaw [11].

REFERENCES

### Table 1.0. Learner’s views towards overall effectiveness of the e-learning

<table>
<thead>
<tr>
<th>Overall effectiveness of the e-learning course</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Maybe/not sure</th>
<th>Disagree</th>
<th>Disagree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>37. The e-learning component adds value to the learning experience for learners.</td>
<td>38%</td>
<td>48%</td>
<td>13%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>38. The e-learning component promotes active learning / problem-based learning / learner centered activities.</td>
<td>34%</td>
<td>47%</td>
<td>14%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>39. I used the e-learning component to support me in my administrative tasks.</td>
<td>18%</td>
<td>27%</td>
<td>45%</td>
<td>9%</td>
<td>0%</td>
</tr>
<tr>
<td>40. I found that the e-learning component supported me in the facilitation of learning.</td>
<td>0%</td>
<td>71%</td>
<td>24%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>41. The e-learning material was easy to use</td>
<td>32%</td>
<td>58%</td>
<td>6%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>42. The e-learning Instructional demos, helped me get started, so I knew what to do when I started exercises/application.</td>
<td>20%</td>
<td>76%</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>43. The step-by-step build up of the demo tutorial solution really helps</td>
<td>21%</td>
<td>68%</td>
<td>4%</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td>44. The step-by-step build up of demo tutorial solution answers makes no difference to my learning style</td>
<td>9%</td>
<td>56%</td>
<td>19%</td>
<td>16%</td>
<td>0%</td>
</tr>
<tr>
<td>45. The step-by-step build up of demo tutorial solution puts me off</td>
<td>0%</td>
<td>60%</td>
<td>8%</td>
<td>32%</td>
<td>0%</td>
</tr>
<tr>
<td>46. The step-by-step build up of demo tutorial solution forces me to work in a way I might not choose.</td>
<td>0%</td>
<td>75%</td>
<td>13%</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>47. I enjoyed practicing the examples along with the multi-media “video” instructions</td>
<td>39%</td>
<td>49%</td>
<td>7%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>48. The e-learning software application is easy to follow</td>
<td>20%</td>
<td>71%</td>
<td>3%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>49. The User-Interface of the e-learning application is easy to understand</td>
<td>19%</td>
<td>71%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>50. There was too much information on the screen , or on the Virtual Learning Environment (VLE)</td>
<td>9%</td>
<td>61%</td>
<td>14%</td>
<td>16%</td>
<td>0%</td>
</tr>
<tr>
<td>51. I appreciate the learner control that's lets me do what I want, as much as I want</td>
<td>7%</td>
<td>71%</td>
<td>18%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>52. I learned to use the CAD/Software application being taught through the e-learning course</td>
<td>26%</td>
<td>59%</td>
<td>13%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>53. I prefer face to face teaching</td>
<td>39%</td>
<td>39%</td>
<td>23%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>54. It is a waste of time doing interactive practice using electronic instructional learning material (i.e. Solid Professor or video posted on VLE; when I can read worked examples from printed manuals.)</td>
<td>0%</td>
<td>46%</td>
<td>8%</td>
<td>46%</td>
<td>0%</td>
</tr>
<tr>
<td>55. I would use e-learning at different stages of my learning, i.e. at basic, intermediate, or advanced levels</td>
<td>20%</td>
<td>69%</td>
<td>10%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>56. The course met my expectations</td>
<td>8%</td>
<td>69%</td>
<td>19%</td>
<td>4%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Effects of Carbon dioxide & Sulphar Contents in Gas Processing Plant

Prof. F. Nabhani, Prof. S. Hodgson & Mr. Abbas M. Jasim
School of Science & Technology
University of Teesside
Middlesbrough TS1 3BA, UK

ABSTRACT

This paper reviews available information in corrosivity prediction in assessment of plant operations, process control and materials selection. The high levels of CO₂ & H₂S with high pressure and temperature will cause corrosion and failures in these units in excess of about $50 million per incident. It is necessary to understand the corrosion impact of associated & refinery off gases in the existing compressor stations, providing safe operating boundaries for various metals at varying hydrogen partial pressure and process temperature in accordance to the applied codes/standards. Given that recommendations in petroleum gas industries for the materials used in CO₂ & H₂S containing environments in oil and gas production. The compressors are addressed broadly, while focus has been retained on ancillary equipment such as scrubbers, coolers and control relief valves. This study also provides broad recommendations for managing any associated corrosion risk such as metal loss, crack integrity assessment, defects in pipeline including its geometric difference, leaks and failure modes. This study was performed in two ways, such as application of NACE MR0175 & EFC 16 Codes / Standards and checking of CO₂/ H₂S / Water corrosion rates.

1. INTRODUCTION

In order to develop a better understanding and prediction tools for CO₂/H₂S corrosion in the future, it is imperative that the basic corrosion mechanisms are well understood. A dedicated study is considered necessary where the basic electrochemical mechanisms of CO₂/H₂S corrosion of bare mild steel will be revisited. This will produce a healthy platform for studying the effect of other complicating parameters, such as formation of protective surface scales, effect of inhibitors, hydrocarbons, glycols, methanol, and condensation in wet gas transport and multiphase flow [1-7].

In this field a low concentration of H₂S (<30 ppm) in a CO₂ saturated water solution can accelerate the corrosion rate significantly in comparison to corrosion in a similar CO₂ environment without H₂S. This “H₂S effect” seemed to vanish at higher H₂S concentrations and higher temperatures (>80°C) when a protective film forms. It was suggested that this effect of H₂S could be significant only in the low pH range (<pH 5) [8-10].

A company has initiated a pipeline network to identify bottlenecks and additional facilities to accommodate a planned increase in through input the facilities from the current 300 MMSCFD to 530 MMSCFD. Refinery Off-gas (ROG) and associated gas (AG) from Petroleum Company is supplied to compressor stations for compression prior to being sent for further processing at Central Gas Plant CGP. The facilities are starting from the gathering lines network at well manifold battery limits, including of additional compressor stations, existing compressor station capacity under projected operating conditions, including compressors and transmission lines, with support from original equipment manufacturers for the turbine compressor packages. For each compressor station, a mixture of associated gas & refinery off gas are received at the inlet scrubber and passed into the compressor suction scrubber. Liquid from the suction scrubber is returned to the GOSP using crude pumps as condensate. The excess liquid carry over beyond the crude pump capacity is drained to a closed drain vessel and the content is routed back to GOSP via crude pumps [11].

This study is to assess the corrosion impact of processing Refinery Off Gas (ROG) mixed with Associated Gas (AG) at compressor stations using NACE National American of Corrosion Engineer MR0175 [12] and EFC 16 European Federation of Corrosion [13] codes/standards. Electronic
Corrosion Engineering (ECE) software was used to determine relative corrosion rates for carbon steel at the most severe process conditions.

2. PROCEDURE

Naturally through the year, the Associated Gas (AG) received from various GOSPs its composition such as Sulphur (H₂S), Carbon dioxide (CO₂); Water content (H₂O) and Hydrogen slightly changes during summer and winter seasons. Table 1 show the composition of AG is mixed with ROG through the inlet line of the compressor station.

<table>
<thead>
<tr>
<th>AG &amp; ROG Composition Elements</th>
<th>Compressor Suction Gas Composition (mole fraction)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Design Case</td>
</tr>
<tr>
<td></td>
<td>AG Summer AG Winter AG &amp; ROG Summer AG &amp; ROG Winter</td>
</tr>
<tr>
<td>H₂S</td>
<td>0.0006</td>
</tr>
<tr>
<td>CO₂</td>
<td>0.0606</td>
</tr>
<tr>
<td>H₂O</td>
<td>0.0619</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1- Compressor Suction Gas Composition

The majority of materials selected during installation of the compressor stations for piping are carbon steel such as API 5L GrA/B and A106 GrA/B. Number of vessels that were used constructed from SA516 Gr70 and there chemical compositions are detailed in table 2. The rest of vessels and piping were constructed from high alloy material such as SS-304 and SS-316.

<table>
<thead>
<tr>
<th>Steel</th>
<th>C</th>
<th>Mn</th>
<th>P</th>
<th>S</th>
<th>Si</th>
<th>Mo</th>
<th>Ni</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>API 5L GrA/B</td>
<td>0.21-0.27</td>
<td>0.90-1.15</td>
<td>0-0.04</td>
<td>0-0.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A106 GrA/B</td>
<td>0.25-0.30</td>
<td>0.27-1.06</td>
<td>0-0.025</td>
<td>0-0.025</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SA516 Gr70</td>
<td>0.27</td>
<td>0.85-1.2</td>
<td>0.035</td>
<td>0.035</td>
<td>0.15-0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SS 304</td>
<td>0.080</td>
<td>2.0</td>
<td>0.045</td>
<td>0.030</td>
<td>1.00</td>
<td>-</td>
<td>9.25</td>
<td>19.00</td>
</tr>
<tr>
<td>SS 316</td>
<td>0.080</td>
<td>2.0</td>
<td>0.045</td>
<td>0.030</td>
<td>1.00</td>
<td>2.50</td>
<td>12.00</td>
<td>17.00</td>
</tr>
</tbody>
</table>

Table 2- Piping & Equipment Chemical Composition

Compression station facility equipment data sheets showing this have been made available by the company and eight streams were selected to reflect the most severe areas which will be affected by the processing condition in the compressor station refer to table 3. These eight stream conditions were examined through the electronic corrosion engineering software to show the effect of pressure, H₂S and CO₂ effects corrosion rate for the following equipment types:

- Inlet, suction and interstage scrubbers
- Fin fan coolers
- Compressor stages
- Piping

In this paper the suction pipeline condition that appears under Inlet Scrubber of the inlet line for the first stream was selected to show the effect of inlet pressure, with related to NACE MR0175 that shown in figure 1 and European Federation of Corrosion EFC 16 figure 2. Gas sourness that is leads to increase in corrosion rate which then calculated electronically through the ECE software. The temperature in the inlet line is considered to be constant; however, fluctuation in pressure, H₂S and CO₂ content varied to establish any change in the corrosion rates. Taking into consideration all the control valves and relief valves are constructed of suitable material, with equal or higher resistance to hydrogen induced cracking than carbon steel.
### TABLE 3 - Process Specifications for the selected streams

**Operating Press [psig]** | 14.996 | 113.631 | 102.631 | 455.296 | 448.296 | 430.996 | 385.231 | 448.296
---|---|---|---|---|---|---|---|---
**Operating Temp [°C]** | 53.437 | 189.061 | 55.556 | 186.378 | 56.111 | 65.556 | 55.982 | 56.111
**Design Press [psig]** | 125.165 | 170.680 | 170.680 | 568.934 | 568.934 | 568.934 | 568.934 | 568.934
**Design Temp [°C]** | 80 | 205 | 85 | 220 | 220 | 95 | 85 | 85
**Line Size inches** | 30 | 16 | 16 | 12 | 10 | 10 | 3 | 1 1/2

**Composition Elements**

| Composition (Mol Fraction) | 0.00056 | 0.00056 | 0.00059 | 0.00059 | 0.00059 | 0.00060 | 0.00039 | 0.00001
|---|---|---|---|---|---|---|---|---
| H₂ | 0.10431 | 0.10431 | 0.10890 | 0.10890 | 0.11140 | 0.10055 | 0.00577 | 0.00003
| Nitrogen | 0.06056 | 0.06056 | 0.06322 | 0.06322 | 0.06322 | 0.06463 | 0.01776 | 0.00048
| CO₂ | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000
| Hydrogen | 0.70840 | 0.70840 | 0.73956 | 0.73956 | 0.73956 | 0.75641 | 0.10055 | 0.00000
| Methane | 0.02421 | 0.02421 | 0.02528 | 0.02528 | 0.02584 | 0.01320 | 0.00000 | 0.00000
| Ethane | 0.01382 | 0.01382 | 0.01443 | 0.01443 | 0.01472 | 0.02055 | 0.00000 | 0.00000
| Propane | 0.00281 | 0.00281 | 0.00294 | 0.00294 | 0.00299 | 0.00868 | 0.00000 | 0.00000
| i-Butane | 0.0230 | 0.0230 | 0.0224 | 0.0224 | 0.0219 | 0.07492 | 0.00000 | 0.00000
| n-Butane | 0.00069 | 0.00069 | 0.00072 | 0.00072 | 0.00072 | 0.00145 | 0.00000 | 0.00000
| i-Pentane | 0.00023 | 0.00023 | 0.00024 | 0.00024 | 0.00024 | 0.00100 | 0.00000 | 0.00000
| n-Pentane | 0.00002 | 0.00002 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000
| n-Heptane | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000
| n-Octane | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000
| C11+ | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000
| H₂O | 0.06191 | 0.06191 | 0.02066 | 0.02066 | 0.02066 | 0.00002 | 0.00185 | 0.99947

### 3. RESULTS

The results obtained from the software are based on five main variables that are partial pressure, temperature, H₂S content and CO₂ content. In this study three of variables were changed such as partial pressure, H₂S content and CO₂ content, the other two such as pipe diameter and temperature were kept constant:

a. With present gas composition as shown in table 3 where CO₂ kept at 6.05 mole%, H₂S at 0.056 mole%, pipe Φ 30" and temperature at 80°C and partial pressure raised to 90 psi the sourness of gas showed in NACE chart only with maximum corrosion rate of 7.9 mpy and pH lowest 4.79 and highest 5.31.

b. Varying in H₂S content two effects occurred in corrosion rate and sourness showed only in EFC 16 chart:

i) Changing operating condition of H₂S to 0.352 mole%, keeping CO₂ as 6.05 mole%, pipe Φ 30" and temperature @ 80°C and reducing the partial pressure to 30 psi as close to operating pressure. The sourness of gas showed up in EFC 16 chart only with maximum corrosion rate of zero mpy and pH of 4.81.
ii) Further change in \( \text{H}_2\text{S} \) to 0.325 mole\%, \( \text{CO}_2 \) to 95.2 mole\%, pipe \( \varnothing 30" \), temperature @ 80\(^\circ\)C and partial pressure 30 psi the sourness of gas showed only in EFC 16 chart with maximum corrosion rate of 35.4 mpy and pH of 4.45.

c. Varying in \( \text{CO}_2 \) content two effects occurred in corrosion rate and sourness showed only in EFC 16 chart:

i) Changing operating condition of \( \text{CO}_2 \) to 6.05 mole\%, keeping \( \text{H}_2\text{S} \) as 1.0 mole\%, in the same pipe \( \varnothing 30" \) and temperature @ 80\(^\circ\)C and partial pressure of 30 psi as close to operating pressure. The sourness of gas showed up in EFC 16 chart with maximum corrosion rate of zero mpy and pH of 4.61.

ii) Further change in \( \text{CO}_2 \) to 100 mole\%, \( \text{H}_2\text{S} \) to 0.0 mole\%, pipe \( \varnothing 30" \), temperature @ 80\(^\circ\)C and partial pressure of 30 psi the gas showed sweet in both NACE and EFC 16 charts; however, the maximum corrosion rate showed with 39.4 mpy and pH of 5.32.

---

**Figure 1 - Sour Gas Systems NACE MR0175**

**Figure 2 - Sour Gas Systems EFC 16**
4. CONCLUSIONS

The conclusion to be drawn from the results obtained above is that the corrosion rate predictor for sweet condition and also for sour service conditions is based on the fundamental chemistry of the solubility of corrosive gasses either $H_2S$ & $CO_2$ contents.

The above results showed with current $H_2S$ content of 0.056 mole%, the total partial pressure can be reached up to about 90 psia to get into sour service refer to NACE MR0175 chart figure 1 above. Since the operating pressure is about 15 psig or even if it reached to 30 psig still with same amount of Sulphar in mole fraction can be operated.

The other element which is found interesting was $CO_2$ content. With the current $CO_2$ level of 6.05 mole% no effect in the corrosion rate; however, with increase of carbon dioxide to 100% then high corrosion rate showed up.

The composition of hydrogen considered during this study was 26 mole %. It is considered that hydrogen content up to 100 % is able to be processed through these compressors without significant risk of hydrogen induced cracking due to the disassociation of molecular hydrogen. It is recommended that the provision of corrosion inhibitor injection be maintained even with the introduction of ROG to compressor stations.

5. REFERENCES

Taguchi Method for Optimising the Sintering Parameter of the Metal Injection Moulding (MIM) Compacts

*K. R. Jamaludin*¹², N. Muhamad¹, M. N. Ab. Rahman¹, S. Y. M. Amin³, S. Ahmad¹³, M.H.I. Ibrahim¹³

¹. Precision Process Research Group
Dept. of Mechanical and Materials Engineering
Faculty of Engineering
National University of Malaysia
43600 Bangi.
Selangor Darul Ehsan. Malaysia

². College of Science and Technology
University of Technology Malaysia
International Campus Kuala Lumpur
54100 Kuala Lumpur. Malaysia

³. Faculty of Mechanical & Manufacturing Eng.
University Tun Hussein Onn,
86400 Batu Pahat, Johor, Malaysia

* Corresponding author: Tel.: (6019) 3862305; Fax: (603) 90745545; E-mail: khairur@citycampus.utm.my

ABSTRACT

Sintering parameters of the SS316L water atomized injection moulded compact has been optimized for its best sintered density. The L₉ (3⁴) Taguchi orthogonal array is used in the experiment while sintering temperature, sintering time, heating rate and cooling rate was selected as factors that influenced the sintered density. The sintering environment was in the vacuum and four replications were done for each trial. The analysis of variance shows that the confident level for the experiment was 99.5 % (α = 0.005) and all factors are highly significant at α = 0.005 to the sintered density. The study concluded that the heating rate has the highest influence to the sintered density (41.29 %) followed by sintering temperature (31.60 %), sintering time (11.13 %) and cooling rate (11.10%). The optimum sintered density obtained is 98.48 % of the theoretical density and the optimum parameter has been verified that the sintered density obtained is in a range of confident interval.

Key words: Sintering; Stainless steel; Taguchi method; Metal injection moulding; Powder injection moulding

1. INTRODUCTION

Metal injection moulding (MIM) is a relatively new processing technology used in powder metallurgy processing industries. This process is especially cost-effective and beneficial for manufacturing small and complex components in large quantities. Metal injection moulding is used in an increasing range of different fields, including automotive, medical and telecommunications industries. It includes four basic steps consisting of mixing the powders and binders, injection moulding, debinding and finally sintering. Both injection moulding and sintering are the most important steps related to forming the green part and the final part respectively.

Therefore, an optimisation of the processing parameter is essential to obtain high quality final part. High sintered density of the final part is vital to maintain an excellent performance of the powder metallurgy products. Many earlier studies about sintering of MIM part [1-5] are concerning with microstructures, densification, and sintering atmosphere. Sintering parameters were optimised by adjusting the sintering variables without using any design of experiment (DOE) methodology. The traditional experimental approach that vary one variable at a time, holding all other variables as fixed does not produce satisfactory results in a wide range of experimental settings. Thus it requires a lot of experiments trial before the optimised sintering parameter is obtained without having any statistical confidence level.
DOE for the injection parameter has been studied by author of reference [6-8] and the results obtained a significant optimum injection parameter for MIM feedstock. As a consequence to the injection parameter optimisation which has been published, this paper presents a sintering parameter optimisation which utilises the published optimised injection parameter. In addition, reference [9] has shown the significance of sintering variables such as heating rate, dwell time, sintering temperature and sintering atmosphere. Reference [9] found that the vacuum environment is best for sintering SS316L and thus, this study attempts to continue the study by using a high vacuum environment in the experiment. Despite cooling rate is replacing the sintering environment in the orthogonal array, heating rate, dwell time and sintering temperature remains as sintering variables for the optimisation. These variables can influence the microstructure, pore size and shape and final density of the sintered parts.

2. EXPERIMENTAL

A MPIF 50 standard tensile bar is used as a specimen. A water atomised 316L stainless steel powder with \(D_{50}\) of 7.157 µm, pycnometer density of 7.90 g/cm\(^3\) is mixed with 73 % PEG weight of polyethylene glycol (PEG) and with a 25 % weight of polymethyl methacrylate (PMMA). In addition about 2 % weight of stearic acid (SA) is used as a surfactant.

Prior to the moulding, compositions are mixed in a sigma blade mixer for 95 minutes at a temperature of 70°C. The greens are prepared by the Battenfeld, BA 250 CDC injection moulding machine while a high vacuum furnace Korea VAC-TEC, VTC 500HTSF with the vacuum pressure of up to \(9.5 \times 10^{-6}\) mbar is used for sintering.

3. DESIGN OF EXPERIMENT

There are many sintering parameters that have some effects on the properties of the sintered density. Therefore, a Design of Experiment (DOE) method is necessary for the experimental work involving a variety of input. The most frequently used methods are partial or full factorial design and the Taguchi approach. With an appropriate DOE, one can quickly and with fewer attempts be able to find out whether these variables have any effects on the quality of the output. The Taguchi approach is mostly used in the industrial environment, but it can also be used for scientific research. The method is based on balanced orthogonal arrays [10]. In this paper, a \(L_9\) (\(3^4\)) orthogonal array consisting of nine experiment trials and four columns is used as a DOE and then followed up by the ANOVA to determine the significant levels and contributions of each variable to the sintered density. The main variables involved in this study are as shown in Table 1. Three levels for each variable refer to the maximum and minimum limit that influences sintered density.

<table>
<thead>
<tr>
<th>Table 1: Factor level (variables) in the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor Level</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
</tbody>
</table>

4. RESULTS & DISCUSSION

The density of the sintered part is measured by the Archimedes immersion method in accordance to the MPIF 42. Four replications are recorded for each experiment as shown in Table 2. The theoretical density shown in Table 2 is calculated from the average of the replication. As shown in Table 2, a combination of \(A_1\), \(B_2\), \(C_0\), \(D_1\) results in the maximum sintered density (98.48 % of the theoretical density) while, a combination of \(A_2\), \(B_0\), \(C_2\), \(D_1\) produces a minimum sintered density (93.53 % of theoretical density). Although the sintering temperature is only 1360 °C, slow heating rate (6 °C/minutes) and a longer dwell time (240 minutes) enable the compact to obtain the maximum sintered density. Nevertheless, with a high sintering temperature (1380 °C) and a quick heating rate at a shorter dwell time will minimise the sintered density.
Taguchi Method for Optimising the Sintering of the Metal Injection Moulding (MIM) Compacts

Besides that, the analysis of variance (ANOVA), demonstrates the significant levels of the variables as well as the effects of the sintering variables on the sintered density as shown in Table 3. Generally, all the sintering variables have significant effects on the sintered density at a 99.5 % significant level. The significant level obtained by this experiment is higher than reported by reference [9]. The ANOVA in Table 3 displays the relative significance of the variables as well as the contributions of the variables assigned to the orthogonal array shown in Table 2. The ANOVA in Table 3 depicts a very significant level (α = 0.005) of each variable. Heating rate (C) is most influential (41.29 %) on the sintered density, followed by the sintering temperature (A), then by the dwell time and finally by the cooling rate. However, reference [9] reported that the sintering atmosphere had been the most significant variable on the sintered density as it demonstrated a much higher variance ratio, F. The sintering atmosphere was the most influential variable (76.685 %) followed by the heating rate (7.377 %), the sintering temperature (5.538%) and finally the dwell time (5.168 %). Thus, based on his study, a high vacuum sintering environment has been considered. Besides that, the influence of the cooling rate on the sintered density is investigated as this variable has not been studied by reference [9] in his DOE. This is due to the fact that the cooling rate is another sintering variable [11]. Although the cooling rate is one of the sintering variables, it has demonstrated to be the lowest in the percentage of contribution. Despite the contribution being low, the high significant level, α as shown by Table 3, still indicates the importance of this variable. This is as important as the dwell time which has been reported to be less influential on the sintered density by reference [9].

With reference to the ANOVA, the main effects of the experiment are calculated by basing on the highest average values as shown in Figure 1. The response plot in Figure 1, shows a combination of A1, B2, C0, and D1 as the highest yield, i.e., where the sintering temperature is at 1360 °C, the dwell time of 240 minutes, the heating rate of 6 °C/minutes and the cooling rate of 8 °C/minutes. On the other hand, a faster heating rate (20 °C/minutes) for the sintering temperature of 1250 °C and a dwell time of 90 minutes has been reported as the optimum sintering parameter by reference [9].

Table 2: Orthogonal array and sintered density

<table>
<thead>
<tr>
<th>Trial</th>
<th>Replication (density(g/cm³))</th>
<th>Variance, ( \mu )</th>
<th>% Theoretical density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: ANOVA for the sintered part at \( \alpha = 0.005 \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Degree of freedom, ( f_n )</th>
<th>Sum squared, ( S_n )</th>
<th>Variance, ( \mu_n )</th>
<th>Pure Sum squared, ( S_n' )</th>
<th>Variance ratio, ( F_n )</th>
<th>Critical F value</th>
<th>Contribution, ( P_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>0.140</td>
<td>0.070045</td>
<td>0.139</td>
<td>114.39</td>
<td>( F_{0.005, 2, 27}=6.4885 )</td>
<td>31.60</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>0.050</td>
<td>0.405058</td>
<td>0.049</td>
<td>40.94</td>
<td>( F_{0.005, 2, 27}=6.4885 )</td>
<td>11.13</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>0.183</td>
<td>0.091317</td>
<td>0.181</td>
<td>149.13</td>
<td>( F_{0.005, 2, 27}=6.4885 )</td>
<td>41.29</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>0.050</td>
<td>0.025001</td>
<td>0.049</td>
<td>40.83</td>
<td>( F_{0.005, 2, 27}=6.4885 )</td>
<td>11.00</td>
</tr>
<tr>
<td>error</td>
<td>27</td>
<td>0.017</td>
<td>0.000612</td>
<td></td>
<td></td>
<td>4.88</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>0.439</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
The ANOVA shown in Table 3 demonstrates that the effects of the variables are significant at 99.5 % significance level. Hence the expected results of optimum performance are as shown in Table 4. The expected optimum performance is as high as 98.48 % of theoretical density while the range of the optimum performance based on 90 % confidence level is 98.22 < \mu < 98.75 % of the theoretical density. The optimum parameter has been proven in the confirmation experiment that is conducted at the combined setting of A_1, B_2, C_0, and D_1 and the results fall within the predicted 90 % confidence interval as shown in Table 4. A density of optimum performance reported by reference [9] is 7.592 g/cm^3, which is lower than that achieved by this study as shown in Table 4.

Table 4: Optimum sintering parameter, optimum performance and confirmation experiment.

<table>
<thead>
<tr>
<th>Optimum parameter:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 B2 C0 D1</td>
<td></td>
</tr>
</tbody>
</table>

(Sintering Temperature, 1360 °C; dwell time, 240 minute; Heating rate, 6 °C/minute; Cooling rate, 8 °C/minute)

Optimum performance: 7.7803 g/cm^3 or 98.48 % theoretical density

Confident interval: ± 0.02 at 90 % confident level (\alpha = 0.1)

Range of optimum performance: 7.7592 g/cm^3 < \mu < 7.8013 g/cm^3 or 98.22 % theoretical density < \mu < 98.75 % theoretical density

<table>
<thead>
<tr>
<th>Confirmation experiment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>g/cm^3</td>
<td>7.8377</td>
<td>7.8365</td>
<td>7.7296</td>
<td>7.7296</td>
<td>7.7834</td>
</tr>
<tr>
<td>% theoretical density</td>
<td>99.21</td>
<td>99.20</td>
<td>97.84</td>
<td>97.84</td>
<td>98.52</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

Sintered density of the water atomised SS316L MIM part is optimised by using the Taguchi method. An L₉ orthogonal array is used to vary the experiment variables. ANOVA shows that all the four sintering variables: sintering temperature, dwell time, heating rate and cooling rate, affect the sintered density significantly. The optimum sintering parameter is found to be A₁, B₂, C₀, and D₁, corresponding to the sintering temperature of 1360 °C, a dwell time of 240 minutes, a heating rate of 6 °C/minute and a cooling rate of 8 °C/minute. Confirmation experiments indicate that when sintering SS316L at an optimum condition, a high theoretical density of 98.52 % can be achieved.

REFERENCES

Modelling the nip in a roll press using circular arc principal stress orientation

A.J. Matchett,
Chemical Engineering, School of Science & Technology, University of Teesside, Middlesbrough, TS1 3BA, England

ABSTRACT
The roll press is used for the compression and compaction of powders. It consists of two, horizontal, co-rotating cylinders with a gap, or nip, between them. Powders, placed on the rolls are drawn into the nip and undergo compaction and agglomeration. Johanson [1] used the principles of bulk solids handling and storage to model the system, and predicted the angle at which the rolls gripped the powder and drew it into the nip. Since then, several other workers have studied the system[2-7]. A problem with the horizontal, incremental slice, force balance approach, used in these studies, is that the stress is not uniform over the slice and the slice does not coincide with the principal stresses, necessitating the presence of shear stresses in the subsequent equations. Enstad showed that an assumption of circular arc principal stress orientation could overcome these problems in hopper design[8]. Matchett [9][10] extended this approach into 2-dimensions and showed that the resultant equations were a co-ordinate specific form of the Lamé-Maxwell equations. This paper presents an application of this approach to the modelling of the nip region. Horizontal and vertical force balances then provide 2 equations from which radial principal stress, $\sigma_R$, and principal arc stress, $\sigma_\epsilon$, may be calculated. Typical data for stress against angle of roll are presented and the results are discussed and compared to the literature. The methodology has potential for development into 2 and 3 dimensions.

INTRODUCTION

A roll press consists of two, contra-rotating cylinders of identical diameter and rate of rotation – Figure 1. Powder is fed into the region above the rolls and is drawn into the nip where it is compressed and granulated.
Johanson[1] provided a rigorous analysis of the system and identified a critical angle, \( \alpha \), at which the material was drawn into the nip: for \( \theta > \alpha \) there is slip between the material and the roll; for \( \theta \leq \alpha \) there is no wall slip and material is drawn along towards the narrowest point of the gap between the rolls at \( \theta = 0 \). Johanson used a horizontal, incremental slice as the basis for his model and related the decrease in volume of the increment in the nip to the compressibility properties of the bulk solid and predicted stresses from the resultant element volume, with no recourse to a force balance.

Kuhn and Downey[2] also used a horizontal increment, but based their analysis upon a force balance. Dec et al[3] presented a similar model. Principal stresses were assumed to align horizontal/vertical throughout the nip. The Kuhn & Downey model only works effectively for an active stress state.

Various workers have compared experimentally determined stresses to the Johanson and Kuhn and Downey models with some success[3-7] . Data are typified by a half of a bell-shaped curve with a significant reduction in the rate of change of normal stress (a stress plateau) as roll angle, \( \theta \), tends to zero – Figure 2.

These models are one-dimensional and it is implicitly assumed that the definition of a horizontal, incremental slice is in some way significant for the averaging of stresses across the element. Unless the horizontal/vertical principal stress orientation is assumed throughout, the element does not align with the principal stresses.

Enstad overcame this problem in an analysis of stresses in hoppers by assuming a circular arc principal stress orientation[8] and defined an incremental element based upon a circular arc. Matchett developed these models further to include 2-dimensional models[9-10].

The present paper will present an analysis of the nip region of the roll press in terms of circular arc principal stress orientation. The model has greater generality than Johanson or Kuhn & Downey, and illustrates the lack of detailed knowledge of stresses within the nip.

**THE MODEL**

Figure 3 shows the elements of the model. A circular arc, AB, of radius R, is coincident with the principal stress direction makes angles \( \beta_r \) and \( \beta_w \) with the normals at the roll and material limit of the arc B. A region of plastic flow of width \( f \) may be incorporated at the limit if the model, and the circular arc assumption extends only to the commencement of the plastic flow zone. The principal stress arc need not be horizontal at this limit-point B, Figure 3. Principal stress \( \sigma_R \) acts normal to the circular arc AB, and principal stress \( \sigma_c \) acts along the arc. \( \sigma_c \) can take different values at the roll: \( \sigma_{cr} \), and at the opposite limit: \( \sigma_{cw} \). These are all innovations from the Johanson and Kuhn & Downey models.
Modelling the nip in a roll press using circular arc principal stress orientation

A detailed view of the incremental element on which the analysis is based is shown in Figure 4. Radial stress $\sigma_R$ acts over horizontal length $h$ and vertical length $h_1$.

Thus, when $f=0$ and $\beta_w=0$, the extreme of the model represents the centre-line between the two rolls and half of the roll nip is modelled.

Furthermore, if $\beta_r=\theta$ the model is a form of the Kuhn & Downey equation.

The presence of the plastic flow zone allows for backflow[1]. This should enable the force balance model[2][3] to be reconciled with the compressibility approach of Johanson[1].
Modelling the nip in a roll press using circular arc principal stress orientation

At the limit of the nip, $\theta=\alpha$, Johanson proposed that $\beta_w=\theta=\alpha$, giving a vertical/horizontal principal stress orientation. He also proposed an active stress state at this point. However, the state of stress played no further part in the analysis, the remainder being purely geometric.

If the limit of the principal stress arc at B is taken as a wall, the model may be used to describe a roll or blade passing over a solid surface, as in a mixer or mill.

Thus, the model has great flexibility.

From the geometry of the system:

\[
\begin{align*}
\text{Arc} &= \beta_w - \beta_r - \theta : \gamma = \frac{\pi}{2} - \frac{\text{Arc}}{2} : R = \frac{h_o + r(1 - \cos \theta)}{2 \cos \gamma \sin (\theta + \gamma - \beta_r)} \\
\omega &= \frac{\pi}{2} + (\beta_r - \theta - \gamma) : \frac{d\beta}{d\theta} = \frac{d}{d\theta} \left[ r \sin \theta - 2 R \cos \gamma \sin \omega \right]
\end{align*}
\]

A vertical and horizontal force balance can be performed over the incremental element in Figure 4, assuming no inertial terms:

\[
\begin{align*}
-\frac{d}{d\theta} [h_{\sigma_R}] - \sigma_{cr} r \cos \beta_r \sin (\beta_r - \theta) + \sigma_{cw} \left( \frac{d\beta}{d\theta} \right) \cos \beta_w \sin \beta_w &= 0 \\
\frac{d}{d\theta} [h_{\sigma_R}] + \sigma_{cw} \left( \frac{d\beta}{d\theta} \right) \cos^2 \beta_w - \sigma_{cr} r \cos \beta_r \cos (\beta_r - \theta) &= 0
\end{align*}
\]

The Johanson and the Kuhn & Downey equations consist of single equations describing stress in the nip. The present model consists of two equations and allows a certain amount of variation of stress across the system in that $\sigma_{cr}$ may not necessarily be equal to $\sigma_{cw}$.

These equations may be integrated numerically to give $\sigma_R$, $\sigma_{cr}$ and $\sigma_{cw}$ as functions of $\theta$. In order to achieve this, appropriate values of the model parameters must be selected.

**RESULTS AND DISCUSSION**

Above the nip region: $\theta>\alpha$, there is wall slip, but while material is often assumed to be at yield, this may not necessarily be the case—there is no over-riding physical reason for this assumption. However, the calculation of stresses in this region is necessary in order to determine the input stress for the nip region: $\sigma_R=\sigma_{Ro}$ at $\theta=\alpha$. This is an extremely sensitive parameter[1][7].

In the nip region, the material is at yield and at the roll:

\[
\frac{|\sigma_R - \sigma_e|}{(\sigma_R + \sigma_e)} = \sin \delta
\]

Values of $\beta$ may be calculated at the limits of wall and material yield as:

\[
\beta_{\text{passive}}, \beta_{\text{active}} = \left( \text{arcsin} \left( \frac{\sin (\phi_w)}{\sin \delta} \right) \right) \pm \phi_w / 2
\]
However, as there is no wall slip within the nip region, it is not possible to use these equations. Experimental data do not provide this information, and usually consist of normal stress data only, as a function of \( \theta \) \[7\]. Therefore, selection of the values of the fitted parameters for the model is difficult on other than purely empirical ground: the model is made to conform to the physical data by an appropriate fitting process.

The first issue is state of stress: active or passive. The Kuhn & Downey model only works with an assumption of active stress: vertical stress is the major principal stress. Johanson\[1\] proposed an active stress state at \( \theta = \alpha \), where the principal stresses take a horizontal/vertical orientation. However, the circular arc model can be used with an active or a passive assumption of stress within the nip. The passive stress state generally gives higher stresses for a given roll gap – Figure 5.

Generally, arc radius \( R \) is positive at \( \theta > \alpha \), but would be negative at \( \theta < \alpha \) if \( \beta_r > 0 \) within the nip region. This implies a “concave outwards” principal stress arc radius, rather than the convex(positive) radius shown in Figure 3 and 4. Thus, \( \theta = \alpha \) may be a switch/transition point.

Selection of an appropriate value of \( \beta_r \) is also problematic and may, in principle, take any value less than the limits given in equation 4.

The presence of the plateau region as \( \theta \) tends to zero may also cause problems-Figure 2. The characteristic plateau is inherent in the Johanson model: the rate of change of volume in the incremental element tends to zero as \( \theta \) tends to zero and Johanson simply calculates the stress associated with the element volume from an empirical, plastic compression algorithm.

In the Kuhn & Downey model, the principal stresses at \( \theta = 0 \) align horizontal/vertical and the plateau comes directly from the resultant equations.

In the more general model presented here, it is necessary that \( \beta_r \) tends to \( 0 \) as \( \theta \) tends to zero in order to describe this feature. A general variation of \( \beta_r \) with \( \theta \) may be proposed, in which \( a \) is a fitted parameter:

\[
\beta_r = \theta + a(\alpha - \theta)
\]

The general effects of some of these parameters can be seen in Figure 5.

**Figure 5: wall normal stresses for system with active and passive stress states.**

Roll radius 0.1m; roll gap .00006mm; \( \delta = 35^\circ \); \( \phi_w = 20^\circ \)

\( \beta_r = 16^\circ \); \( \beta_w = 0 \) except where stated. The passive case refers to the right-hand axis.
Figure 5 shows passive stress giving rise to higher stresses than the active state, for equivalent conditions. In both cases, with a constant value of $\beta$, there is a lack of a plateau as $\theta$ tends to zero- hence to need to use equation 5. The presence of a flow region, typified by $\beta_w$ not equal to zero, changes the stresses. For $\beta_w>0$ ($\beta_w=16^\circ$) this causes a relaxation in stresses.

The present model can be made to conform well to physical data. Typical data of Bindhumadhavan et al.[7] are shown in Figure 6. These data have been modelled assuming no backflow: $f=0$ and symmetry about the centre-line: $\beta_w=0$. Values of fitted parameters have been adjusted to give a fit to the data, including: $\sigma_R$; $\alpha$; state of stress; $\alpha$. This gives the model great flexibility, but reflects a lack of detailed knowledge of stress conditions within the nip region.

The model and data presented represent an initial study of the roll nip, using circular arc principal stress orientation methods. There is considerable scope for further work, including:

- Utilisation of the model’s flexibility to deal with a wider range of situations
- Reconciliation of the force balance model with Johanson’s plastic compressibility model, using the facility for backflow in the plastic flow zone
- Development of the model into a 2-dimensional system to investigate stress variations across the nip-see[10]

However, there is a lack of detailed experimental data in order to validate this and other models. Thus, at the macro-level, there are several models which fit available experimental data, but all are lacking precision at the more detailed level.

**CONCLUSIONS**

A model of stresses within the nip region of a roll press has been presented. The model contains several innovations compared to previous one-dimensional analyses. However, there is a lack of knowledge for the selection of appropriate parameters to fit the model to data.

The model is very flexible and may be applied to a number of other situations involving rolls and/or blades in the processing of powders.
REFERENCES


NOTATION

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arc</td>
<td>span of circular arc</td>
<td>[rad]</td>
</tr>
<tr>
<td>dB/dθ</td>
<td>rate of change of location of point B with θ</td>
<td>[m/rad]</td>
</tr>
<tr>
<td>h</td>
<td>horizontal length of action of stress σR</td>
<td>[m]</td>
</tr>
<tr>
<td>h₀</td>
<td>value of h at the roll-gap</td>
<td>[m]</td>
</tr>
<tr>
<td>h₁</td>
<td>vertical length of action of stress σR</td>
<td>[m]</td>
</tr>
<tr>
<td>R</td>
<td>circular arc radius</td>
<td>[m]</td>
</tr>
<tr>
<td>r</td>
<td>roll radius</td>
<td>[m]</td>
</tr>
<tr>
<td>α</td>
<td>critical value of nip angle θ</td>
<td>[rad]</td>
</tr>
<tr>
<td>β₀</td>
<td>wall normal angle at roll</td>
<td>[rad]</td>
</tr>
<tr>
<td>β₀w</td>
<td>wall normal angle at centre or flow zone</td>
<td>[rad]</td>
</tr>
<tr>
<td>γ</td>
<td>angle=π/2−Arc/2</td>
<td>[rad]</td>
</tr>
<tr>
<td>θ</td>
<td>roll normal angle</td>
<td>[rad]</td>
</tr>
<tr>
<td>σR</td>
<td>radial stress acting on circular arc</td>
<td>[Pa]</td>
</tr>
<tr>
<td>σ₀</td>
<td>value of σR at θ=α</td>
<td>[Pa]</td>
</tr>
<tr>
<td>σ₀r</td>
<td>arc stress acting at roll</td>
<td>[Pa]</td>
</tr>
<tr>
<td>σ₀cw</td>
<td>arc stress acting at centre line of flow zone</td>
<td>[Pa]</td>
</tr>
</tbody>
</table>
Use of Mechatronic System to Manufacture Spiral Bevel Gear using 3-Axis CNC Milling Machine

Mohsen Safavi, Mahsa Ghandehary*, Reza Abedinzadeh, Mehdi Karimian
Islamic Azad University, Najaf Abad Branch
Isfahan, Iran

ABSTRACT
CNC machining nowadays makes more and more use of "Mechatronics", in other words combining numerical control with mechanic, electric and data processing systems can lead to new methods of production. In recent years, the development of CNC has made it possible to perform nonlinear correction motions for the cutting of spiral bevel gears. Gears are crucial components for modern precision machinery as a means for the power transmission mechanism. Due to their complexity and unique characteristics, gears have been designed and manufactured by a special type of machine tools, such as gear hobbing and shaping machines. The spiral bevel gears (SBG) are the most complex type and are used to transmit the rotational motion between angularly crossed shafts. In this paper, we attempt to manufacture the spiral bevel gear by a three-axis CNC milling machine interfaced with an additional PLC module based on traditional discontinuous multicutting method accomplished by using a universal milling machine interfaced with an indexing work head. This research consists of (a) Geometric modeling of the spiral bevel gear (b) Simulating the traditional and our new non-traditional method using a CAD/CAE system (c) Process planning for CNC machining and PLC Programming (d) Experimental cuts with a three-axis CNC milling machine were made to discover the validity of the presented method.

1. INTRODUCTION
Gears are important and precision mechanisms for industrial machinery as a means for mechanical power or motion transmission between parallel, intersecting and non-intersecting cross-axis shafts. Although hidden from sight, gears are one of the most important mechanical elements in our civilization. They operate at almost unlimited speeds under a wide variety of conditions. The machines and processes that have been developed for producing gears are among the most ingenious we have. Whether produced in large or small quantities, in cell, or job shop batches, the sequence of processes for gear manufacturing requires four sets of operations [1]:

1. Blanking
2. Gear cutting
3. Heat treatment
4. Grinding

Dependent on their type and application or required strength and resistance, gears are manufactured by casting, extruding, forging, powder metallurgy, plastic molding, gear rolling and machining. Among these processes, machining is more frequently used for high precise gears [2]. Among the various types of gears (Figure 1), the spiral bevel gears (SBG) are the most complex type and are used to transmit the rotational motion between angularly crossed shafts. Gears that have teeth curved longitudinally along the length of the teeth. The main advantage of these gears over the straight-toothed varieties lies in the fact that as more teeth are in contact at the same time because of the curved-shaped contour of the teeth, a smoother meshing action between the mating pair is ensured [3].

* Corresponding author: E-mail: ghandehary@yahoo.com
Figure 1: Various industrial gears
Spiral bevel gear is a term generally used in case of bevel gears.

The design and manufacturing of spiral bevel gears is still a hot topic of research that is vital for application of such gears in helicopter transmissions, motorcycle gears, reducers, and in other branches of industry. As far as manufacturing is concerned, the gears are machined by a special type of machine tools, such as gear hobbing and shaping machines. Recently, special CNC based gear manufacturing machine tools are used in industrial practice. Figure 2 shows one of those special machines used to manufacture Spiral Bevel Gears. This may be why literature on gear manufacturing is sparse in the open research domain. Recently, CNC based gear manufacturing machine tools have been developed and increasingly used in industrial practice. However, their kinematic structure is still inherently different from the industrial CNC milling machine, as the former is designed for a special type of cutter.

Figure 2: Special machine tool and cutter for manufacturing SBGs

In this paper, we attempt to present a new manufacturing procedure of SBGs by using a three-axis milling machine interfaced with a PLC module which operates as an indexing table. In terms of production rate, it is obvious that this method will be lower than that of the special machine tool. Other than production rate, this method is advantageous in the following respects: (1) the conventional method requires a large investment for obtaining various kinds of special machinery and cutters dedicated to a very limited class of gears in terms of gear type, size, and geometry; (2) by this method, various types of gears can be manufactured with the industrial three-axis CNC milling machine; (3) this method is more economical than using the special machine tool. In [4], Suh and Jih presented a complicated mathematical tool path planning (called tool path planning for sculptured surface machining...
(SSM) of SBGs) for a ball endmill which is used for cutting the surface model of the spiral bevel gears based on basic gearing kinematics and involute geometry along with tangent planes.

2. Geometric specifications of the Spiral Bevel Gears

Most of the time, the geometric parameters of a gear are provided with an engineering drawing. Figure 3 shows Spiral-Bevel-Gear nomenclature. Some parameters (principal parameters) are required for defining the geometry. Table 1 summarizes these parameters including relationships among parameters for the gear which has been manufactured in our test. To calculate these parameters we have used “drive component development software” called GearTrax.

![Figure 3: Spiral-Bevel-Gear nomenclature](image)

Table 1: Gear Data

<table>
<thead>
<tr>
<th>Spiral Bevel Gear Specification</th>
<th>Pinion</th>
<th>Gear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module</td>
<td>m=3</td>
<td></td>
</tr>
<tr>
<td>Pressure Angle</td>
<td>( \Phi = 20 )</td>
<td></td>
</tr>
<tr>
<td>Spiral Angle</td>
<td>( \psi = 35^\circ )</td>
<td></td>
</tr>
<tr>
<td>Shaft Angle</td>
<td>( \delta = 90^\circ )</td>
<td></td>
</tr>
<tr>
<td>No. of Teeth</td>
<td>( Z_1 = 20 )</td>
<td>( Z_2 = 20 )</td>
</tr>
<tr>
<td>Spiral direction</td>
<td>Left Hand</td>
<td>Right Hand</td>
</tr>
<tr>
<td>Face Width</td>
<td>( b = 19 )</td>
<td></td>
</tr>
<tr>
<td>Pitch Angle</td>
<td>( \alpha_1 = 45 )</td>
<td>( \alpha_2 = 45 )</td>
</tr>
<tr>
<td>Root Angle</td>
<td>( S_{1_1} = 40.802 )</td>
<td>( S_{1_2} = 40.802 )</td>
</tr>
<tr>
<td>Face Angle of Blank</td>
<td>( S_{1_1} = 49.198 )</td>
<td>( S_{1_2} = 49.198 )</td>
</tr>
<tr>
<td>Whole Depth</td>
<td>( K = 5.6640 )</td>
<td></td>
</tr>
</tbody>
</table>

The design of spiral bevel gear requires high-accuracy mathematical calculations and the generation of such gears drives require not only high-quality equipment and tools for manufacturing of such gear drives, but the...
development of the proper machine-tool settings. Such settings are not standardized, but have to be determined for each case of design (depending on geometric parameters of the gear drive and generating tools) to guarantee the required quality of the gear drives.

Figure 4, illustrates model of spiral bevel gear created by GearTrax.

3. MANUFACTURING THE SBG

As it is discussed in the introduction, by machining, all types of gears can be made in all sizes and machining is still unsurpassed for gears that most have very high accuracy. Form milling is one of the most common machining processes used to manufacture any types of gears. The cutter has the same form as the space between adjacent teeth. Standard cutters usually are employed in form-cutting gears. In the United States, these come in eight sizes for each diametral pitch and will cut gears having the number of teeth indicated in standard tables.

To manufacture the SBGs with the 3-axis CNC milling machine, we first test the process by developing a CAD/CAM system composed of geometric modeling and graphic simulation modules. The commercial software SolidWorks is used for creating CAD model and Visual Nastran is used for simulating the process of gear manufacturing (Figure 5).

As far as machine tool configuration is concerned, it is obvious that a rotational motion of the workpiece is required for NC machining of the SBGs. Based on the machinability analysis [5], at least four-axis controls are required for NC machining of SBGs by one set-up. Thus, a rotary table to be interfaced with the three-axis milling machine is required. Form cutting or form milling is used in our tests. The tool is fed radially toward the center of the gear blank to the desired tooth depth, then across the tooth face, while the rotary table rotates the workpiece
around its center, to obtain the required tooth width. When one tooth space has been completed, the tool is withdrawn, the gear blank is indexed using a dividing head, and the next tooth space is cut. Basically, this method is a simple and flexible method of machining SBGs. The equipment and cutters required are relatively simple, and standard 3-axis CNC milling machine is used. However, considerable care is required on the part of tool feed which should be a small value in each step to prevent any spoil.

Although the form cutting of this kind of gears is currently done on universal milling machine, using an indexing head, the process is slow and requires skilled labor and operator. Figure 6 shows the arrangement that is employed when the work is done on a universal machine in Isfahan-Ferez workshop. As it is shown, the cutter is mounted on an arbor, and a dividing head is used to revolve (required to cut the gear tooth) and index the gear blank. The table is set at an angle equals to the spiral angle (35 degrees), and the dividing head is geared to the longitudinal feed screw of the table so that the gear blank will rotate as it moves longitudinally. In the presented method, we have used an AC motor interfaced with a worm gearbox. Worm gearbox is used to reduce the output speed of AC motor and also to set the angle between the tooth trace and the element of the pitch cone, known as spiral angle [6].

Because of the needed synchronization between tool path planning and rotary motion of outward shaft of worm gearbox is required, we used a mechatronic system to control all the four axes (one-axis motion for the rotary table and three-axis motion for the cutting tool) simultaneously.

![Figure 6: A photograph of manual cutting on universal milling machine](image)

This mechatronics system contains five major parts: (1) PLC (Programmable Logic Controller), one of electronic equipments, which can be called “Sequence Controller”. It uses internal microprocessor to execute the calculation of sequence of the work according to the status and the value of the input signal of Encoder and Sensors; (2) Inverter, which is used to operate a standard 3-phase motor and act as an intermediate device between PLC and AC motor. It has different input terminals which can be used to rotate AC motor in Forward/reverse direction; (3) Rotary Encoder, a feedback system which determine the position of outward shaft or measure the revolution of workpiece mounted on outward shaft; (4) Proximity sensors, which produce an output signal if a metal object (form cutter in our experiment) enters the electromagnetic field of its oscillator. Two proximity sensors are placed on the system where define the start and end of the tool path. (5) A photoelectric sensor, a device used to detect the presence of an object by using a light transmitter which consists of a receiver located within the line-of-sight of the transmitter. The power supply has two states of light on mode and dark on mode. In first state, the control output turns on by receiving light and turns off by interrupting light. In second state, the control output turns off by receiving light and turns on by interrupting light. In this mode, an object is detected when the light beam is blocked from getting to the receiver from the transmitter. These sensors are mounted on a ring around the outward shaft with a longitudinal crack through it. We used ladder diagram, common program language, to operate the PLC in the mechatronic system. The operation of the PLC based on the ladder diagram is as follows: Step 1: Read the external input signal, such as the status of sensors or rotary encoder. Step 2: Calculate output signal, according to the value of the input
signal in the step 1 and send it to AC drive (Inverter) to run the AC motor in Forward/reverse direction or turn the motor for a special angle (Circular pitch) using a rotary encoder.

The procedure of the whole system is accomplished in five stages;

Stage 1: Form cutter reaches the first proximity sensor and sends a signal to the PLC. As it’s mentioned before, PLC sends out an output signal to the AC drive to run the motor in forward direction.

Stage 2: Form cutter machines the rotating workpiece.

Stage 3: Cutting tool reaches the second proximity sensor and sends the second signal to the PLC. PLC sends a stop command to the inverter.

Stage 4: Milling tool withdraws the stopped workpiece and returns to its first place.

Simultaneously, AC drive runs the motor in backward direction until the shaft reaches the first position and the receiver of photoelectric sensor sees the transmitted radiation passes through the longitudinal crack on the shaft.

Stage 5: Stage 1 through stage 4 continues till the first tooth space is cut. PLC counts a number for each of the above four stages till it reaches the predefined number of machining sequences. Then it sends out a signal to the AC drive to index the gear blank equal to diametral pitch and all the above stages will be repeated again. The diametral pitch is measured by a 1024-pulses/revolution rotary encoder.

Figure 7: A photograph of electrical panel including PLC and Inverter

The traditional procedure (Figure 6) is quite slow, and considerable care is required by the operator. The procedure utilized is essentially the same as on a universal milling machine, except that, after setup, the various operations are completed automatically. The machining configuration and a photocopy of the experimental cutting are shown in Figures 7 and 8 and 9.

Figure 8: A photograph of the experimental cutting
4. SUMMARY

In this paper we attempted to manufacture the spiral bevel gear using a three-axis CNC milling machine based on form milling method. Basically, form cutting uses a simple and flexible method of machining of gears. The equipment and cutters required are relatively simple and inexpensive, and standard CNC milling machine is used. Compared to the conventional gear cutting method in which dedicated machine tools is required, the presented method can easily be modified to produce any type and size of SBGs or any other types of gears.

Also it is a multi-deception system (use of mechatronic and CNC systems) which is a dynamic and constant evolving technology.

Having looked at all the mentioned stages above, the following diagram can be considered as an overall picture of this invention:
In an advantageous embodiment of the method according to the present invention, machining time is one of the most factors which have been greatly noted. For example, it took only 2 minutes to cut one tooth space completely. While, in traditional method, it took more than half an hour to cut same tooth space.

In a further advantageous embodiment of the present invention with respect to manual cutting, the instant for the angular compensation of the cutter head is set shortly before the end of the plunge process.

The workpiece is wood and the blank material is pre-machined as a conic (Face Angle of the gear) form by turning operation. Standard cutter No.5 used in our experiment is mounted on the machine spindle and the gear blank is mounted on outward shaft of worm gearbox. The tool is fed subsequently (around 30 machining sequences to prevent spoiling work) toward the center of the gear blank to the desired tooth depth. When one tooth space has been completed, the tool is withdrawn; the gear blank is indexed using the AC motor based on explained procedure, and then is followed by cutting the next tooth space.

REFERENCES

Aspects of Burr Formation in Bandsaw Teeth Manufactured by Milling Operation

Mohammed Sarwar* and Julfikar Haider

School of Computing Engineering and Information Sciences
Northumbria University
Newcastle upon Tyne, NE1 8ST, UK

ABSTRACT

Steel manufacturers and stockholders prefer bandsawing for cutting off bar, tube, pipe etc. compared to other techniques as it enjoys competitive advantages of higher accuracy of cut, better surface finish, lower kerf loss, better straightness of cut, long tool life, high metal removal rate etc. Along with the geometries of the bandsaw tooth (e.g., back to tip height, root to tip height, rake angle, set magnitudes, set balance etc.), the bandsaw cutting edge condition (e.g., edge sharpness, burr etc.) significantly affects the cutting performance of a bandsaw. Currently the production of bandsaw is largely done by milling operation due to the scale of manufacturing and the economics of milling compared to other processes (e.g., grinding). Ideally, the bandsaw teeth should possess sharp cutting edges with no burr. In general, two types of burr are commonly seen in the bandsaw teeth manufactured by milling operation namely tooth tip burr and side burr. Current research undertaken at Northumbria University in collaboration with a major bandsaw producer is focused on the mechanism of burr formation in the bandsaw teeth. The paper briefly outlines the factors affecting the burr formation in bimetal (HSS edge wire and soft steel backing material) bandsaw teeth manufactured by milling process and suggests the necessary steps to be considered for manufacturing burr free bandsaw with sharp cutting edges. The investigation showed that flank wear in the milling cutter has a major influence on the side burr, whereas tooth tip burr was influenced by both flank wear and “V” type notch wear found at the crossover point on the flank face. It was also observed that TiN coating on the milling cutter could control the burr formation in bandsaw teeth to some degree.

1. INTRODUCTION

Technically, burr in machining process is defined as an undesirable projected material generated on the edge of a machined workpiece as a result of plastic deformation. Burr could be generated in any machining process such as turning, milling, drilling, grinding etc.. Burr formation on the edges of machined components causes various problems in manufacturing. For example, if a machined part goes through various manufacturing steps, the presence of burrs on the edges of the part causes misfit problem in subsequent assembly operation hence, strongly affecting the productivity. In automated production environment, burr could give erroneous measurement due to dimensional inaccuracy, thus affecting product quality. The performance of a final precision product could degrade due to the burr. Therefore, in most of the cases, costly (up to 30% of the total product price [1]) and time-consuming deburring process need to be incorporated in the production line. Furthermore, in some cases conventional deburring techniques might not be suitable to eliminate the burr completely. Consequently, burrs represent a serious bottleneck in machining operations in terms of productivity, product quality and production cost. Thus, urgent attention is necessary to control the burr formation in the machined parts, thereby saving production cost and time.

Material removal by milling process is considered as one of the most multipurpose and extensively used machining processes. Theoretical and experimental studies have been performed to understand the mechanism of burr formation on machined parts in milling operation [2-6]. During the milling operation, burrs are created mainly where the milling cutter exits the machined part. Burr generation in bandsaw teeth produced by milling operation is also a serious concern because the presence of burr on the bandsaw cutting edges significantly affects the cutting performance. A clear understanding of the burr formation mechanism would be great value in minimising the burr
and hence, improving the manufacturing precision of bandsaw teeth. The aim of this paper is to provide valuable insight for better understanding of the burr formation in bandsaw teeth produced by milling operation.

2. **BANDSAW MANUFACTURING PROCESS**

The manufacturing process of bandsaw involves the packing of long bimetal strips in a milling machine and forming the teeth in the band strips using a specially designed milling cutter (Figure 1). The milled bandsaw teeth are set in left and right hand directions using different set patterns and then hardened to give the required hardness and toughness in the bandsaws.

![Figure 1: Bimetal band strip, milling cutter, milling machine and generated bandsaw teeth in the strip](image)

The cutting tool in this case is a form milling cutter in which the flutes are arranged in a spiral manner. Sharp bandsaw teeth are formed by the complex cutting action of the two teeth in successive flutes of the milling cutter (Figure 2). Interrupted cutting action takes place in milling operation. As the workpiece material is not a solid block, rather a stack of many thin strips, the interrupted cutting characteristics are even more pronounced in the milling process. The milling of bimetal strips even complicates the machining process as the milling cutter encounters different materials (HSS, welding and spring steel) along the depth of cut.

![Figure 2: A schematic diagram of bandsaw teeth generation by a milling cutter](image)

3. **MATERIALS, CUTTER AND MILLING MACHINE**

The workpiece used for manufacturing bandsaws is long bimetal steel strips, in which M42 HSS edge wire is welded to D6A backing material as shown in Figure 1. Powder Metallurgy High Speed Steel (PM HSS; 900 HV) is used for the solid milling cutter, which is hardened and tempered at different stages of manufacturing of the cutter to give the maximum strength at the cutting edges. Bimetal bandsaws are manufactured in CNC milling machines (Figure 1) under wet machining condition. Bandsaw cutting edges and milling cutter wear were examined in a
Compact Video Microscope (Allen CVM; magnification 50×) and optical microscope. The bandsaw tooth quality (Burr, back to tip height etc.) was measured in a Shadowgraph (BATY International Ltd. P14 XLQC) having a magnification of 25×.

4. TYPES OF BURR IN BANDSAW TEETH AND ITS CONSEQUENCES

Researches on burr formation in milling process [2-4] have identified different kinds of burr depending on the type of operation (face milling, end milling or slot milling), shape of burr and location of burr in machined workpiece. Machining bandsaw teeth by milling process is a unique milling operation where two distinct kinds of burr are generally observed namely, side burr and tooth tip burr. Side burr generally appears around the gullet and along the cutting edges; whereas tooth tip burr appears in the bandsaw tooth tip (Figure 3). The existence of burr is sometime hard to recognise by naked eyes, but it is clearly appeared under microscope or in shadowgraph. Side burr can be felt by the fingers when rubbing the machined edge.

Milled bandsaws are subsequently subjected to setting operation to set the bandsaw teeth in opposite directions. The set magnitudes are continuously measured from the top by an online camera system. The presence of side burr gives higher readings hence, interrupts the production process and affects the quality of bandsaws. The tooth tip burr also results an increase in back to tip height of bandsaw, which could also affect the set magnitudes.

At the new condition, edge radius of a milled bandsaw tooth is approximately 20-50 μm [7], which is in many cases equal to the depth of cut in bandsawing operation. Figure 4 illustrates the difference in chip formation mechanisms by a lathe tool and by a bandsaw tooth. The presence of tooth tip burr alters the cutting edge condition and the burr breaks as soon as it engages with the workpiece to be cut. As a result, the cutting edge radius increases and could become higher than the depth of cut. In this situation, sliding and ploughing processes dominate instead of the cutting process [5, 7] and consequently, bandsaw cutting performance decreases.
5. DISCUSSIONS ON BURR FORMATION

5.1. BURR FORMATION MECHANISM

Figure 5 presents a schematic diagram of a pack of band strips to be milled in order to produce bandsaw teeth. The milling cutter starts forming teeth at the back band first and then progresses forward to form teeth in all the bands. Exit side burr, which is the most common type of burr, is generated in the bandsaw due to the plastic flow of material at the exit of the milling cutter teeth. The mode of machining operation is generally down/climb milling, where the cutter exits with a zero chip thickness. Therefore, at the end of a cut instead of actual shearing, plastic deformation process dominates and the deformed material is pushed down underneath the cutter. This encourages the generation of side burr in bandsaw teeth [4]. It has been suggested in the literature by avoiding the exit of the cutting edge of milling cutter, exit burr formation can be minimised [6]. Exit side burr in bandsaw teeth is minimised by exiting the cutting edges of the milling cutter in the burr plate and fixed jaw. Side burr is more prominent in the back bands as it is more close to the exit of the cutter and only gets support from the burr plate and fixed jaw. On the other hand, least or no burr at all appears in the front band as it has the support of all the previous bands in the pack plus the burr plate and fixed jaw. In any particular band, no side burr appears at the front side of a band where the cutter enters. Tooth tip burr is generated at the bandsaw tooth tip due to the scissoring action of the milling cutter teeth.

![Figure 5: Schematic illustration of milling process showing vice jaws, band pack, burr plate and the cutter feed direction](image)

5.2. FACTORS INFLUENCING BURR FORMATION

Burr formation in the workpiece machined by milling operation is dependent on a combination of factors including workpiece materials, machining mode of operation, cutter geometry, cutter wear and machining parameters [2, 4, 6]. Gillespie [8] pointed out that tool sharpness is one of the most significant parameters affecting the size of burr produced in milling. At the start of a milling operation very little or no side burr appears in the bandsaw teeth as the cutting edges of the milling cutter is sharp. After machining considerable number of milling passes, the sharpness of milling cutter edges deteriorates due to the development of flank wear. The lack of edge sharpness in the cutting edges alter the chip formation mechanism where the ploughing process becomes dominant instead of shearing and consequently, the level of burr generation in bandsaw teeth increases. Therefore, the burr formation (both side burr and tooth tip burr) is directly proportional to the cutter wear. Other researchers also reached the same conclusion about the relationship between burr formation and cutter wear during machining stainless steel [5].

Figure 6 shows an example of tooth tip burr generation in bandsaw teeth and its relation to cutter wear. The higher back to tip height in a particular bandsaw at the end of a milling operation compared to that at the beginning.
could be related to the tooth tip burr. The development of flank wear in the cutter edges at the end of the milling operation could explain the reason for tooth tip burr formation (Figure 7).

Figure 6: Back to tip height of bandsaws in a milling pass at the beginning and end (after several hundred passes) of a milling operation. Averages of 5 passes are presented.

Author’s earlier work [9] established that TiN coating on milling cutter reduced the flank wear hence, contributing to the minimisation of burr formation in bandsaw to some degree. Olverai and Barrow also pointed out that coated milling cutter produced slightly smaller burr than an uncoated one when milling carbon steel [4]. It has also been established that generation of sharp bandsaw teeth is responsible for a “V” type notch wear at the crossover point on the flank face in some milling cutters. Excessive notch wear could be related to tooth tip burr (Figure 2). It was evidenced that TiN coating could reduce the notch wear and hence, a reduction of tooth tip burr [9]. Figure 8 shows the development of notch wear in coated and uncoated teeth under identical machining condition.

Burr plate plays a central role in reducing the side burr formation in the back bands of the pack. Burr plates are periodically fed in the upward direction to support the bandsaw teeth. Burr plates could be jammed in the burr plate guides due to small dimensional inaccuracies (e.g., non-uniform width along the length) in the burr plate. This causes increased level side burr in the back bands as the band pack gets insufficient support from the burr plate. Special care should be taken when setting the burr plate in the guides and adjusting the amount of feed so that burr plate can provide continuous support on the milled teeth in the pack. Figure 9 shows an example where it can be
seen that due to the poor setting of burr plate, one side of the burr plate moves higher than the other side (Figure 9b) hence, the band pack gets lack of support from the burr plate and side burr generates in the back band.

Figure 8: Notch wear generated on the flank face of the milling cutter teeth

Figure 9: (a) Correct movement (straight upward) and (b) incorrect movement (twisted) of the burr plates

The band pack width (total milled workpiece) varies depending on the size of the bandsaw tooth forms to be milled. It has been observed that side burr is more prominent in the bands milled in a wider band pack than those milled in a narrower band pack. This can be explained by the fact that under certain vice pressure wider band pack could be less stable and subsequently a higher degree of side burr appears in the bands. Poor condition of vice and vice jaws could also lead to burr formation due to lack of support on the band pack.

6. SUMMARY

The paper presents the core ideas of bandsaw teeth formation by milling process and the generation of burr in the bandsaw teeth. It has been found that side burr and tooth tip burr are the two most common type of burr in the bandsaws teeth. The side burr is related to the flank wear in the milling cutter and the tooth tip burr is related to both flank and notch wear on the flank face. The tooth tip burr changes the edge condition (edge sharpness, edge radius, edge shape etc.) of the bandsaw teeth, which could be detrimental to the bandsaw cutting performance. Therefore, it is absolutely vital to maintain a sharp cutting edge in the milling cutter teeth to avoid or to minimise burr in bandsaw teeth. TiN coating has positive influence in minimising the burr formation through the reduction of flank and notch wear in the milling cutter. Other influencing factors identified are dimensional accuracy of burr plate and guides, setting of burr plate in the guides, vice jaw condition, vice pressure etc..

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support from the Bandsaw Company and the School of Computing, Engineering and Information Sciences, Northumbria University, UK to carry out this investigation.
Aspects of Burr Formation in Bandsaw Teeth Manufactured by Milling Operation

REFERENCES

Cutting Energy and Wear in Bandsawing Operation When Cutting Ti-6Al-4V Alloy

Fahd N. Khan¹⁺, Mohammed Sarwar¹ and Martin Persson²

¹School of Computing Engineering and Information Sciences
Northumbria University
Newcastle upon Tyne, NE1 8ST, UK

²SNA Europe (AB)
R&D Department
Box 833, SE-531
Lidköping, Sweden

ABSTRACT
Titanium and its alloys have found wide application in the aerospace, automotive and biomedical industries owing to their good strength to weight ratio and excellent corrosion resistance. However, these alloys have poor machinability, which is attributed to their high strength and low thermal conductivity leading to high cutting temperatures. This paper reports on the experimental data associated with the forces, and specific cutting energy with the wear of tungsten carbide tipped (TCT) bandsaw blades while machining titanium alloy Ti-6Al-4V.

1. INTRODUCTION
Bandsawing is a widely used metal cutting operation in a variety of industries, cutting-of to size raw material for secondary processes. This cutting-off process can offer an advantage of high automation possibilities, low kerf loss, straightness of cut, good surface finish and a long tool life [1]. Bandsawing is now a well understood process, due to the scientific work carried out by several researchers [2-5], which was stimulated by demands for higher efficiency, better accuracy and improved surface quality.

The characteristic feature of the material removal in bandsawing operation is the function of the cutting edge with limited sharpness (5 µm to 15 µm) with the layer of material being removed also being very small (5 µm to 50 µm). The cutting action in bandsawing operation is intermittent, with several cutting edges in contact with the workpiece material. This is one of the major differences between the sawing operation and the other single-point cutting operation, e.g. turning, where only one sharp edge is in contact with the workpiece material. Moreover, the chip needs to be accommodated in the gullet and ejected at the end of the cut [6].

Titanium and its alloys are used extensively in aerospace owing to their excellent combination of high specific strength (strength-to-weight ratio), which is maintained at elevated temperature, their fracture resistant characteristics, and their exceptional resistance to corrosion. They are also being used increasingly (or being considered for use) in other industrial and commercial applications, such as petroleum refining, chemical processing, surgical implantation, pollution control, nuclear waste storage, food processing, and marine applications [7]. They have become established engineering materials available in a range of alloys and in all the wrought forms, such as billet, bar, plate, sheet, strip, hollows, extrusions and wire. The machinability of titanium and its alloys is generally considered to be poor owing to several inherent properties of the materials. Titanium is chemically reactive and, therefore, has a tendency to weld to the cutting tool during machining, thus leading to chipping and premature tool failure. Its low thermal conductivity increases the temperature at the tool/workpiece interface, which affects the tool life adversely. Additionally, its high strength maintained at elevated temperature and its low modulus of elasticity further impairs its machinability.

Traditionally, high speed steels (HSS) and cemented carbides have been employed to cut/machine titanium alloys. The main disadvantage of high speed steel cutting tools is that it undergoes severe plastic deformation when cutting at elevated temperatures above 600 °C. Tungsten carbide cutting tools have proven their superiority in...
almost all the machining process and interrupted cutting (end milling, tapping and broaching) of these difficult to cut alloys. Technical feasibility of further improvement of cutting properties and life of tool materials by design of new compositions has been essentially exhausted [8]. That is why a great deal of attention has been directed towards improvement of life and wear resistance of the metal cutting tools by producing hard surface coatings.

One of the challenges in design of cemented tungsten carbide tools is the optimization of toughness with straight tungsten carbide alloys with the superior crater resistance of alloyed carbides containing high levels of titanium carbide. This has led to the development of coated carbide tools, which amounts for the major portion of all commercial metal cutting inserts sold worldwide.

Different kinds of coatings, physical vapour deposition (PVD) or chemical vapour deposition (CVD), and different coating materials (e.g. TiC, TiN, TiCN, Al2O3) have led to their application in specialized areas [9]. Both are used to deposit single and multilayer coatings. In PVD coatings, the deposition temperature is below 600 °C and therefore does not adversely affect the bulk mechanical properties. The use of CVD techniques where the temperatures are in the range of 850-1000 °C means that the substrate material will have to be re-treated to restore the bulk mechanical properties [10]. Moreover, with PVD coatings, the sharpness of the cutting edge is greater, making cutting forces smaller and hence lengthening tool life.

Previous work [6] was related to the bandsawing of steels, using bi-metal bandsaws. This paper reports the initial work of the present research, which is focused on use of un-coated and coated tungsten carbide bandsaws for machining titanium alloy Ti-6Al-4V.

2. CUTTING TEST PROGRAMME

Full product testing is expensive and time-consuming. In the current research work, the performance of the un-coated and coated bandsaw teeth was evaluated using a “single tooth time compression technique” previously developed by Sarwar [11].

2.1. CARBIDE TEETH DETAILS

All the teeth were examined for any defects, and geometrical features were measured using optical microscope. The details for the uncoated and coated teeth are presented in Table 1.

Table 1: Measured and tabulated properties of the un-coated tooth and TiN coated tooth

<table>
<thead>
<tr>
<th></th>
<th>Uncoated tooth</th>
<th>Coated tooth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. tooth tip thickness</td>
<td>1.599 mm</td>
<td>1.590 mm</td>
</tr>
<tr>
<td>Rake angle</td>
<td>9.8 ° (avg)</td>
<td>9.9 ° (avg)</td>
</tr>
<tr>
<td>Clearance angle</td>
<td>32.1 ° (avg)</td>
<td>32.2 ° (avg)</td>
</tr>
<tr>
<td>Tooth damage</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Appearance</td>
<td></td>
<td>Golden yellow</td>
</tr>
<tr>
<td>Cutting edge radius</td>
<td>7 µm (avg)</td>
<td>7 µm (avg)</td>
</tr>
<tr>
<td>Microstructure of edge</td>
<td>Tungsten carbide, cobalt binder</td>
<td>Tungsten carbide, cobalt binder</td>
</tr>
<tr>
<td>Microstructure of backing material</td>
<td>Tempered martensite</td>
<td>Tempered martensite</td>
</tr>
</tbody>
</table>

2.2. SURFACE TREATMENT OF THE TEETH

A batch of teeth was coated specifically for these tests. They were titanium-nitride coated by Oerlikon Balzers, using the PVD arc evaporation technique, giving the coating a thickness of approximately 5 microns. The chemical composition of the coated teeth is shown in Table 2.

Table 2: Chemical composition of Balzers TiN coated TCT (Tungsten Carbide Tipped) tooth, measured using the EDX. Average of 3 TiN coated teeth

<table>
<thead>
<tr>
<th>Element</th>
<th>Titanium</th>
<th>Nitrogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Wt. %</td>
<td>82.34</td>
<td>17.66</td>
</tr>
</tbody>
</table>
2.3. Workpiece Material Composition and Hardness

Workpiece material (Ti-6Al-4V) chemical composition and mechanical properties are shown in Table 3 and Table 4 respectively.

Table 3: Chemical composition of Ti-6Al-4V. Measured using the EDX system attached to scanning electron microscope

<table>
<thead>
<tr>
<th>Element</th>
<th>N</th>
<th>C</th>
<th>Fe</th>
<th>O</th>
<th>Al</th>
<th>V</th>
<th>Ti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. wt%</td>
<td>0.05</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
<td>60</td>
<td>4.0</td>
<td>Bal.</td>
</tr>
</tbody>
</table>

Table 4: Tensile and mechanical properties of uni-directionally rolled Ti-6Al-4V sheet [12]

<table>
<thead>
<tr>
<th></th>
<th>Tensile strength (MPa)</th>
<th>Yield strength/elongation</th>
<th>Tensile modulus (GPa)</th>
<th>Hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal</td>
<td>945</td>
<td>870</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>Transverse</td>
<td>1105</td>
<td>1061</td>
<td>7.5</td>
<td>130</td>
</tr>
</tbody>
</table>

2.4. Cutting Conditions

Two different feed rates were chosen for the cutting tests, using both the un-coated and coated teeth. These cutting conditions are shown in Table 5.

Table 5: Machining parameters for the cutting tests

<table>
<thead>
<tr>
<th></th>
<th>Cutting speed (m/min)</th>
<th>Feed (µm)</th>
<th>Avg. width of cut (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set A</td>
<td>50</td>
<td>30</td>
<td>1.402</td>
</tr>
<tr>
<td>Set B</td>
<td>50</td>
<td>50</td>
<td>1.402</td>
</tr>
</tbody>
</table>

3. Results and Discussions

3.1. SEM Analysis of the New and Un-used Teeth

In order to observe the cutting edges of the worn teeth, all the 4 teeth were observed under scanning electron microscope before any tests were carried out. All these are shown in Figure 1.

The edge is smooth and appears to be free of any defects on both the rake and clearance faces. The coating is continuous and smooth on both the faces.

3.2. SEM Analysis of the Teeth after 800 Cuts

The un-coated and coated teeth after performing 800 cuts were observed under the electron microscope in order to observe the wear and the degradation that took place. The images are shown in Figure 2.

The images show that the cutting edges are intact for both the un-coated and coated teeth, although they seem to be wearing out at one of the corners. This effect is more pronounced in the un-coated teeth compared to the coated teeth.
Figure 1: SEM images of the un-used un-coated and coated teeth, showing the morphology of the unused surface.
3.3. SEM ANALYSIS OF THE TEETH AFTER 1500 CUTS

All the teeth (un-coated and coated) after performing 1500 cuts were observed under the electron microscope in order to observe the wear and the degradation that took place. The images are shown in Figure 3. It is evident from the images that the un-coated teeth have experienced a greater degree of wear as compared to the coated teeth. This degradation can be clearly observed in the images shown in Figure 4. The corners of the un-coated teeth have experienced a great degree of wear and degradation, and small particles of the workpiece material seem to adhere to the surface of the un-coated teeth, suggesting that the mechanism of wear is adhesive wear. Although, workpiece particles are also present on the surfaces of the coated teeth, but they are less in number as compared to the particles adhering to the un-coated teeth.
Figure 3: SEM images of the un-coated and coated teeth after 1500 cuts, showing the morphology of the cutting edges and the faces.

Un-coated Sample A; Depth of cut 30 µm; 1500 cuts

Un-coated Sample B; Depth of cut 50 µm; 1500 cuts

TiN coated Sample A; Depth of cut 30 µm; 1500 cuts

TiN coated Sample B; Depth of cut 50 µm; 1500 cuts

Figure 4: SEM images of the un-coated and TiN-coated teeth, after 1500 cuts, showing the wear at the corners of the edges.

Un-coated sample
Depth of cut: 50 µm
1500 cuts

TiN coated sample
Depth of cut: 50 µm
1500 cuts

Corner wear

Corner wear
3.4. Specific Cutting Energy Determination

The specific cutting energy has been calculated using the following equation:

\[ E_{sp} = \frac{F_v}{A_{\text{chip}}} \]  

Where, \( E_{sp} \) = specific cutting energy (J/m\(^3\)), \( F_v \) = cutting force (N) and \( A_{\text{chip}} \) = cross-sectional area of undeformed chip (m\(^2\)).

The \( E_{sp} \)-value is a measure of the energy required to remove a unit volume of workpiece and will reflect the efficiency of the cutting process.

The specific cutting energy for all the cutting tests were measured and are plotted against the total length of workpiece material machined. The results for the un-coated and coated teeth at 30 µm and 50 µm depth of cut are presented in Figure 5 and Figure 6 respectively.

![Figure 5: Variation in \( E_{sp} \) with the length of cut for the un-coated and coated teeth. Depth of cut: 30 µm, average width of cut: 1.402 mm](image)

![Figure 6: Variation in \( E_{sp} \) with the length of cut for the un-coated and coated teeth. Depth of cut: 50 µm, average width of cut: 1.402 mm](image)
The above results show that TiN-coated bandsaw teeth to perform more efficiently based on the specific cutting energy values. The lower specific cutting energy is due to the combination of the following:

- Immediate formation of smooth and continuous chip as soon as the cutting edge comes into contact with the workpiece material.
- The flow of the chip along the rake face without any hindrance, which is due to the improved friction conditions created at the tool/workpiece interface because of the TiN coating.

The reduction in the Esp clearly indicates the improvement in cutting conditions and the cutting efficiency when cutting with TiN-coated teeth. It is also evident from the energy values as well as from the SEM images that there was less adhesion of the workpiece material on the tool, because of the TiN coatings.

4. CONCLUSIONS

- The PVD TiN coatings can be applied to the segments of the bandsaw blades, used for machining titanium alloys.
- Simulation cutting tests have proved that TiN coatings improve cutting performance and life, as demonstrated by the improvement in specific cutting energies.

REFERENCES

Thermoplastic Starch/LLDPE Blends as a Binder in Metal Injection Molding

Norhamidi Muhamad* & Hooman Abolhasani
Department of Mechanical and Materials Engineering
Universiti Kebangsaan Malaysia
Bangi, Selangor, 43600, Malaysia

ABSTRACT

Metal injection molding (MIM) is a technique to produce near net shape component combining the injection molding productivity and the strength of powder metallurgy. Binders in MIM are composition of polymers which provide flow possibility of metal powders through the molds as well as shape retention of particles during debinding process. Development of environmentally friendly, low cost and readily available binder systems has received great interest of researchers in this field. In this study, a new environmentally friendly binder has been developed for (MIM) process. Tapioca starch as a nontoxic and water soluble natural polymer has been introduced as the base component of the binder. Linear low density polyethylene (LLDPE), glycerol, citric acid and stearic acid have been blended with starch to make this binder system. Before mixing starch with LLDPE, plasticization process was carried out on starch in presence of glycerol in order to make starch flowable. Citric acid was utilized as compatibilizer of plasticized starch and LLDPE. 316L stainless steel water atomized powder in 20 µm average particle size was utilized to make a feedstock with 57% vol. powder loading. Rheological properties of binder and feedstock have been investigated through a CFT-500D Shimadzu capillary rheometer. Thermogravimetric analysis on feedstock was conducted in order to understand decomposition behavior of binder components in mixture of powder and binder. Viscosity values obtained by capillary rheometry for binder and feedstock settle in suitable range (below 10 Pa.s for binder and below 1000 Pa.s for feedstock). Correlation of viscosity and temperature for the binder and feedstock were determined from flow curves. Rheological characteristic of feedstock exhibited pseudo-plastic or shear thinning flow behavior that verifies implementation of this new binder system for MIM process.

1. INTRODUCTION

Metal injection molding is a newly developed metal forming technology for producing small, complex and precision parts. The research on binder development has been the core study of this method. In metal injection molding, the binder formulation consists of three main ingredients: a backbone which keeps the injected articles in place during debinding stage, the filler which is removed in debinding stage to open pores for burn out remained binder in sintering phase, and finally the surfactant, a polymer which provides adherence between binder and powder [1].

Starch as a biodegradable polymer has been widely used in different industries. The ability of starch in mixing with synthetic polymers makes it behave similar to conventional polymers [2]. Among several characteristics of starch, being environmentally friendly and its wide availability and low cost makes it an outstanding candidate for specific applications. Starch granules uptake water at temperatures higher than 50°C and provide a colloidal gel [3]. Addition of plasticizer can improve the ductility of gelatinized starch. This latter combination is called thermoplastic starch (TPS) [4].

In this study, blend of thermoplastic starch and linear low density polyethylene (LLDPE) has been used as a binder in metal injection molding (MIM) process. Thermoplastic starch and LLDPE are immiscible, thus an additional factor should be used to make a blend of them. Among several TPS/LLDPE compatibilizers available are maleic anhydride, ethylene vinyl alcohol and citric acid. Citric acid was chosen due to a simple process required to blend TPS and LLDPE [5].

* Corresponding author: Tel.: (006) 3 8921 6500; Fax: (006) 3 8925 9659; E-mail: Hamidi@eng.ukm.my
2. METHODOLOGY

2.1. MATERIALS

Native tapioca starch was kindly provided from National Starch and Chemicals (M) Sdn Bhd. Citric acid anhydrous (Batch No. 070215FGKUH and M=192.13 g/mol), glycerol anhydrous (Batch No. 071121KUHUSP and M=92.10 g/mol) and stearic acid was purchased from SYSTERM Co. Linear Low Density Polyethylene (LLDPE) supplied by TITANEX with melt flow index=25 g/10min and batch No. LI2516. Characteristics of binder components are given in Table 1.

<table>
<thead>
<tr>
<th>Material</th>
<th>( \rho ) (g/cm(^3))</th>
<th>Melting point (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native Tapioca Starch</td>
<td>1.5</td>
<td>250</td>
</tr>
<tr>
<td>Glycerol</td>
<td>1.261</td>
<td>18</td>
</tr>
<tr>
<td>Citric Acid</td>
<td>1.665</td>
<td>153</td>
</tr>
<tr>
<td>Stearic Acid</td>
<td>0.847</td>
<td>69.6</td>
</tr>
<tr>
<td>LLDPE</td>
<td>0.918</td>
<td>130</td>
</tr>
</tbody>
</table>

The metal powder used for feedstocks preparation in these experiments was water atomized 316 L stainless steel powder with a 20 µm particle average size which was purchased from EPSON ATMIX Corporation.

2.2. STARCH PLASTICIZATION

At the first stage, glycerol as a usual starch plasticizer in presence of distilled water was applied to tapioca starch plasticization. Mixture of tapioca starch, distilled water and glycerol were blended in a sigma-blade mixer. In this study a 36% glycerol content was utilized for samples preparation. According to Rodriguez Gonzalez et al. (2004), this fraction was proven as the best composition of plasticized starch. In this stage, excess water content in thermoplastic starch evaporates due to the working point temperature [4]. Finally a yellow transparent fluid was appeared that was very sticky and could be considered as a thermoplastic starch indeed.

2.3. THERMOPLASTIC STARCH/LLDPE BLENDING

Thermoplastic starch was heated up to 130°C (melting point of LLDPE) in mixer and LLDPE and citric acid were added to thermoplastic starch simultaneously. Citric acid was added as a starch-LLDPE compatibilizer to make a white and homogeneous paste.

2.4. BINDER PREPARATION

Blend of TPS and LLDPE cannot be mixed with metal powder and therefore, another constituent is necessary to complete the binder formulation. As mentioned before, surfactant agent is able to wet the surface of powder particles and join the binder and powder to make a homogenous paste. Stearic acid as a common surface agent was added to melted TPS/LLDPE blend. Cooling of the recent paste resulted in a stiff and adequate hard solid. Finally, this hardened stack generated crumbs was crushed for the subsequent processes. Table 2 shows the composition of the binder.

<table>
<thead>
<tr>
<th>Material</th>
<th>wt.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starch</td>
<td>41.2</td>
</tr>
<tr>
<td>Glycerol</td>
<td>23.4</td>
</tr>
<tr>
<td>LLDPE</td>
<td>28.5</td>
</tr>
<tr>
<td>Citric Acid</td>
<td>1.9</td>
</tr>
<tr>
<td>Stearic Acid</td>
<td>5</td>
</tr>
</tbody>
</table>
2.5. **Feedstock Preparation**

A mixture of binder and metal powder was prepared to obtain a feedstock with 57% vol. powder loading [6]. Blend of thermoplastic starch and LLDPE was melted in mixer at 130°C and afterward mixture of metal powder and stearic acid was gradually added to provide a homogeneous paste.

**Material Characterizations**

Rheological behavior of prepared samples was assessed using a CTF-500D Shimadzu Capillary Rheometer. The samples were extruded through a die with 1 mm diameter and 10 mm length. A series of experiments were performed to investigate the rheological behavior of the thermoplastic starch, binder and the prepared feedstock. The rheological behavior of the thermoplastic starch was carried out at three different temperatures which were 70, 90 and 105°C, while this attribute of binder evaluated at temperatures 120, 130 and 140°C. Based on the obtained results, three temperature sets of 160, 170 and 180°C were selected to analyze the flow behavior of the 57% vol. feedstock.

Thermogravimetric analysis performed on feedstock in order to determine decomposition characteristics of binder ingredients. In this test, 48 mg sample was selected to observe thermal decomposition behavior between 35 to 600°C.

3. **RESULT & DISCUSSION**

The flow characteristic of thermoplastic starch was evaluated through capillary rheometer at three temperature points including 70, 90 and 105°C. The measurements indicated the viscosity, $\eta$, versus shear rate, $\dot{\gamma}$. The result of these experiments is demonstrated in Figure 1.

![Figure 1: Relationship between shear rate and viscosity of thermoplastic starch](image)

Thermoplastic starch exhibits a pseudo-plastic or shear thinning behavior which means by increasing shear rate, the viscosity values decreases. Hence, TPS satisfies the basic requirement of flow characteristics for a good binder component in MIM.

Rheological behavior of binder was investigated at 120, 130 and 140°C. The derived results are illustrated in Figure 2.
This binder not only exhibits a pseudo-plastic behavior, but also its viscosity values are less than 10 Pa.s. This attributes verifies this binder as a suitable binder for MIM process.

Figure 3 illustrates the result of rheological measurements of the feedstock. These experiments were carried out at three ranges of temperature including 160, 170 and 180°C German and Bose (1997).

It has been stated by German and Bose (1997), during injection molding, shear rate fluctuates between 100 and 100000 s\(^{-1}\). In this range, the suitable values of viscosity for binder are lower than 10 Pa.s meanwhile, this value for feedstock should be less than 1000 Pa.s. Rheological behavior of feedstock with 57% powder loading satisfies this initial requirement of a good feedstock for MIM.

According to general equation of shear rate dependency of viscosity is given by;

\[ \eta = K\dot{\gamma}^{n-1} \]  

(1)
where $\eta$ is viscosity, $K$ is a constant, $\dot{\gamma}$ is shear rate and $n$ is flow behavior index. The values of $K$ and $n$ can be calculated from the graph of viscosity versus shear rate. The gradient of $\log \eta$-$\log \dot{\gamma}$ is equal to $n$-1, and then flow behavior index readily could be obtained. Table 3 shows the $n$ values for binder and feedstock at different temperatures.

Table 3: Flow behavior index

<table>
<thead>
<tr>
<th>Composition</th>
<th>Temperature (°C)</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binder</td>
<td>120</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.45</td>
</tr>
<tr>
<td>Feedstock</td>
<td>160</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>0.40</td>
</tr>
</tbody>
</table>

The $n$ values less than 1 imply the pseudo-plasticity of binder and feedstock. Apparently, the pseudo-plasticity of binder is more than feedstock due to powder content of feedstock. By increasing temperature in feedstock this property increases.

One of the most important aspects in injection molding is flow activation energy ($E$) which indicates correlation of viscosity to temperature. According to Arrhenius equation,

$$
\eta = \eta_0 \exp \left( \frac{E}{RT} \right)
$$

This parameter can be derived by drawing $\ln \eta$ versus $1/T \times 10^3$ at specific shear rates. In this equation $\eta_0$ is viscosity at reference temperature, $E$ is flow activation energy, $R$ is gas constant and $T$ is temperature.

Figure 4 shows the graphs which have been derived by the described method for binder at different shear rates including 1000, 3162.3 and 10000 s$^{-1}$.

In order to find out relationship between temperature and viscosity for feedstock, the graph of $\ln \eta$ versus $1/T \times 10^3$ at various shear rates consist of 398.1, 1000 and 2511.9 s$^{-1}$ in Figure 5 has been plotted.
Figure 5: Temperature dependency of viscosity for feedstock (A: $\gamma = 398.1 \text{ s}^{-1}$, B: $\gamma = 1000 \text{ s}^{-1}$ and C: $\gamma = 2511.9 \text{ s}^{-1}$)

The calculation of flow activation energy can be derived through the equations representing the plotted lines. These values for binder and feedstock have been determined and are given in Table 4.

Table 4: The influence of shear rate on the temperature dependency of viscosity

<table>
<thead>
<tr>
<th>Composition</th>
<th>Shear rate ($\text{s}^{-1}$)</th>
<th>$E$ (KJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binder</td>
<td>1000</td>
<td>29.2</td>
</tr>
<tr>
<td></td>
<td>3162.3</td>
<td>29.8</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>30.5</td>
</tr>
<tr>
<td>Feedstock</td>
<td>398.1</td>
<td>18.51</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>22.01</td>
</tr>
<tr>
<td></td>
<td>2511.9</td>
<td>25.51</td>
</tr>
</tbody>
</table>

Interestingly, the feedstock needs lower flow activation energy rather than binder at the same shear rate ($\gamma = 1000 \text{ s}^{-1}$). The achieved results for flow activation energy of feedstock are in agreement with other works, which has been mention in references [7], [8] and [9]. As the mentioned works are done on binders which are established in MIM process, this agreement confirms the validity of the obtained results for a starch based binder.

The results of TGA on feedstock with 57% vol. powder loading are shown in Figure 6. This graph represents pyrolysis process of the multi-component binder.
It can be observed that the TGA profile is generally divided into three major sections. The first major mass loss is 2.81%, which starts at 160°C. This decomposition is related to glycerol. Starch pyrolysis occurred at interval of 270 to 392°C, which totally comprise 3.37% mass loss. Thermal degradation of LLDPE took place at temperatures above 400°C with totally 3.05% mass loss.

Obviously, mixing and injection molding temperature must be set lower than the minor decomposition temperature. In this study the temperatures above 200°C will results in the degradation of the binder components.

4. CONCLUSION

The results of this study reveal the pseudo-plastic behavior of thermoplastic starch, starch based binder and prepared feedstock. Furthermore, viscosity values of binder and feedstocks exist within the suitable range for metal injection molding process. Thermo gravimetric analysis determines the mixing and injection molding thermal limitations and on the other hand debinding process schedule as well. This indicates the starch can be used as a binder component in MIM process.

ACKNOWLEDGEMENT

The authors would like to thank the Malaysian Ministry of Science, Technology and Innovation for sponsoring this work under project 03-01-02-SF0206.

REFERENCE


On Commercial off the Shelf Design of Reconfigurable Machines Tools

K. Mpofu 1*, C. M Kumile 1, and N. S Tlale 2

1 Department of Mechanical Engineering, Tshwane University of Technology, Pretoria, South Africa
2 Department of Material Science and Manufacturing Centre for Scientific and Industrial Research, Pretoria, South Africa

ABSTRACT

This work seeks to address the challenges met by manufacturing which is based on fixed machining systems or slightly flexible machining systems. Reconfigurable machining systems are identified as the solution to this challenge. An overview of the methods that have been used in the conception of reconfigurable tools in the past is also given. The development of models of the reconfigurable tools utilising the commercial off the shelf (COTS) approach is herein considered. Structures of the attained machines and their different applications are presented. Aspects which resolve the challenges seen in the previous manufacturing epochs are clearly highlighted.

1. INTRODUCTION

Reconfigurable Manufacturing Systems (RMS) embrace a new order of machine tool design that seek to address the inadequacy and insufficiency of the previous epochs in manufacturing. The two major trends of production in the past have been the dedicated manufacturing systems (DMS) and flexible manufacturing systems (FMS) which are now inappropriate to meet the current demands of the global economy. The RMS paradigm is rooted in making manufacturing cost effective, achieving high quality and rapid responsiveness. This is achieved by combining the advantages of the previous epochs and eliminating their disadvantages [1].

The characteristic of rapid responsiveness is the epicentre of the reconfigurable paradigm. This is now very possible particularly with the modular approach that these systems are designed to follow [2]. The availability of modular components as Computer Aided Design (CAD) models has also been a springboard to the realisation of these machine tools using the commercial off the shelf modules from a wide range of suppliers [3].

Using both the functional description of machine tools and the structural description of machine tools authors determine in this work to establish a methodology for the configuration of these machine tools. These machines are designed to be based on commercially available modules from CAD model suppliers on the web. It is envisaged that the critical trait of rapid responsiveness will be achieved, for the more global market adequate manufacturing firm. The responsiveness will be in terms of variation in requirements in both lot sizes and variants of products. In this work authors review the range of machine tool structural configurations in a bid to establish a commercial of the shelf method for structural configuration of the reconfigurable machine tools (RMTs). The rest of the paper is organised as follows; the second section is a literature survey, sections briefly consider the approach utilised in the research. The fourth section elaborates on the COTS implementation, preliminary concepts are then discussed and a conclusion of the paper is provided.

* Corresponding author: Tel.: (0027) 12 382 5286; Fax: (0027) 382 5022; E-mail: mpofuk@tut.ac.za
2. Literature Survey

In this section the different approaches to the machine tool design are considered. There is a wide disparity in the views held by institutions; however to some degree institutes attain to some similarity in the design of their adaptable machine tools.

2.1 Machine Design Methods

A method called the future oriented machine tool (FOMT) design technique has been proposed by Narita et al. [6]. It consisted of four function blocks (management, prediction, observation and strategy) which are regarded as the required abilities which a human operator has when they endeavour to execute the manufacturing task and are given in summary below;

- **Management**: This is the block that manages and prepares the machining know-how data; this includes the process data, resource data, constituting the accumulated data which becomes the planning data.
- **Prediction**: This block aids in the monitoring of the status of the machine before the process begins, during the process and after the process.
- **Strategy**: This block decides the machining method using the data from the management and prediction block. Information is also manipulated in a real-time manner also from the observation block.
- **Observation**: With the utilisation of sensors signals intelligent monitoring is realized to identify machining problems. The information from the other blocks is stored for future utilisation.

Shinno and Ito utilise a variant design method in their establishment of different configurations for machine tools [9, 10].

At the University of Michigan strides have been taken in establishing a way of constituting reconfigurable machine tools. Screw theory was used here to single out components from a database of precompiled modules having been given the respective machine tool requirements [5].
Butala and Sluga [8] viewed the architecture of a machine tool as a system structure which was reflected in its configuration and which impacted the systems performance. The interfaces of the system where depicted by interface between a process and other working system elements. They constructed their understanding of the system around the ‘elementary work system’ as a reasonable representation of the architectural issues of the reconfigurable machine tool.

As a process implementation device (PID) in a machining work system they articulate a machine tool, composed of three subsystems;
1. The Positioning subsystem- between the tool and the work piece
2. The Kinematic subsystem- providing relative process motions in terms of the cutting speed and the feed rate.
3. Energy subsystem- Delivers required energy for machining.

These subsystems, which may be viewed as the major components, are complimented by the minor subsystems such as the following:
1. The interface between the tool and the machine.
2. The interface between the work-piece and the machine.
3. The interface converting coded references to control signals.
4. The amplifiers and converter of control signals.
5. The parameters enabling closed loop control.

Butala and Sluga [8] described the cutting process view whereby each surface generating machining process can be classified into translation and rotation process movements. This approach was also implemented in a computer aided planning system, to clarify the need of having the embedded features to be implemented in the collective drives that constitute it. This resulted in an adaptation of the methodology for the design of reconfigurable assembly systems.

3. UTILISED DESIGN APPROACH

The architectural development of the industrial machine tool used was the COTS modules being used in conjunction with customised developed modules for the completion of a machine tool. This method builds the machine tool based on the commercially available functional units and thereafter completes the tools with modules that are not available ‘off shelf’. There is a wide range of approaches in designing machine tools the researchers consider a hybrid solution generated from the different approaches (with a minimal amount of customised modules).

3.1 COTS APPROACH
In order to implement the COTS method, the following steps are considered;

- A decision is made on the machine tool that is required.
- The machine tool is broken down into its respective constituent modules or sub systems.
- Search for the individual modules from the database of commercial off shelf modules, which has been developed from the Internet.
  - The library is created by reflecting on the needs available in the industries of concern.
  - The modules are classified for easy retrieval from the database
- Establish different configurations of machine tools satisfying the stated requirements.

4. COMMERCIAL OFF SHELF MODULE IMPLEMENTATION

Machine design has conventionally followed a methodology where by a range of concepts are developed then one is chosen for further development and refinement. In the design of reconfigurable machine tools the same trend has been taken by a significant number of researchers. Having realized that the subsystems of the machine are also available as discrete entities, they can also be gathered to create a database or library of modules which can be used to assemble a modular machine as per customer requirement. In the development of products there are mainly two types of products, the integral product development and the modular product development. A distinct definition is given by Stone, Wood and Crawford [4] who clarified the disparity of product types as follows “An integral architecture is defined as a physical structure where the functional elements map to a single or a number of very
small physical elements”. On the other hand “Modular architectures are physical product sub-structures that have a one-to-one correspondence with a subset of a product’s functional model”.

4.1 Why Off-Shelf Modular Approach

- The normal route in design of machine tools has been to design the concepts of the final machine then break it down to its respective modules. Modularity has often been an afterthought. However if modularity is identified at the initial conceptual or reverse engineering effort the immediate product design reaps benefits in reduced development time and costs.
- Modular machines have been available however they have not been designed to be reconfigurable. An advantage that can be harnessed from changing the configuration of the machine tool is the consequent reaping of a long lifespan of the modules.
- There is a need for a new body of knowledge in machine tool design with respect to reconfigurability and modularity. This knowledge base would be realised through better productivity of the manufacturing sector and its competitiveness in meeting rapidly changing global market needs.
- The production cost of the modular machine will be lower as the time in developing new concepts is reduced. As will be realized in the deliberations later on, there are some basic modules that will virtually not change and only a selected portion of the entire module assembly has to be designed and fitted to meet new customer requirements. Thus the cost of designing the entire machine tool will be less costly.
- Standardization will be encouraged in both the local and international machine tool industry, in a bid to establish module integratability.

5 Preliminary Conceptual Designs

This section describes the different configurations of machine tools that have been modelled and the functions of these machines structures are briefly highlighted.

5.1 Gantry Type Machine Tool

- The Gantry Type machine tool is suitable for the machining of relatively long parts.
- The machine displays a basic three DOF structure but may be upgraded to a four DOF structure through the integration of additional library modules.
- The axes ranges of the tool may be increased or reduced through the interchange of modules from the mechanical library.
- This machine tool architecture is feasible for drilling, milling, tapping, vertical boring, grinding and polishing applications through the interchange of a machining heads.

![Figure 2 Family of Gantry Type Machine](image)

5.2 Multi Table Machining Platform

- The platform has three linear slides giving two axis motions in the X-Y plane and a third vertical motion.
The rotary axis is responsible for the orientation of the tool relative to the four work tables.
The machine is capable of machining four parts in sequence, located on four worktables.
This machine is capable of milling and drilling functions.

Figure 3 Conceptual Multi Table Machining Platform

5.3 Lightweight Turning Structure

- The design constitutes of a gantry with two linear axes.
- A rotary head positioned horizontally provides rotary motion via a chuck.
- The tool is attached to a linear slide balanced by a vertical column and the cutting motion of the tool.

Figure 4 Lightweight Turning Structure

6 Conclusion

The challenge of achieving reconfigurability at machine level for RMS has been addressed by this research. The realization that the creation of a single machine cannot address the problem of machining systems for RMS has resulted in the succession of a methodology. This method may be applied universally for reconfigurable machine tool development. The tactic outlines how a machine tool may be developed from commercially available mechanical modules being assembled in conjunction with customized user developed modules. The resultant class of machines, named Modular Reconfigurable Machine tools, is able to display a variation of cutting degrees of freedom, a variation of machining functions and the ability to reconfigure axis ranges on a single machining platform. It is these characteristics that will enable RMS machining systems to be reconfigured to meet changing production requirements. For the practical testing of this methodology a reconfigurable vertical machining center has been earmarked for construction. Complimentary to the mechanical conception of the machine tool, a conceptual modular distributed control architecture will be considered for implementation. The hardware of the electronic control system will be fully modular in nature, allowing the system to be customized while still possessing the ability to be upgraded and extended if additional processing potential is required.
To complete the system a model for development an Open Architecture Control system will be considered. This will render the conceptual design complete with all key elements being at a theoretical level, thereby forming the platform for the core engineering and practical development of the Modular Reconfigurable Machine.

ACKNOWLEDGEMENTS

Authors acknowledge the funding and opportunity provided by Council of Scientific and Industrial Research, Advanced Manufacturing Technology Strategy and the Tshwane University of Technology.

REFERENCES

The introduction of a KPI to monitor Spent Foundry Sand disposal

Martin. G. McKie*1, Farhad Nabhani1, Kirk Bridgewood1 and Richard. D. McKie1

1School of Science and Technology
University of Teesside
Middlesbrough TS1 3BA, UK

ABSTRACT

Many Large Companies and Small to Medium Enterprises (SME’s) in the United Kingdom are being forced to change the way they manage and dispose of waste. This change has been caused in part by an increase in landfill taxes introduced by the Government to encourage companies large and small to dispose of waste more responsibly via recycling. By recycling the companies can help reduce the waste taken to landfill, in turn increasing a companies’ sustainability and reducing the environmental effects of disposing of waste at landfill.

This research examines a method of applying a (KPI) Key Performance Indicator to monitor the spent foundry sand disposed of over a one year period in an SME. This monitoring of the KPI then initiated changes within the company to find alternative methods of reducing the spent foundry sand taken to landfill. There were a number of methods examined which could potentially offer new ways for the SME to recycle the spent foundry sand. Some of the various new methods examined were, for example, using the sand in cement and concrete production as well as the possibility of using the sand in sand bags for flood defence in heavily flooded areas of the UK. This paper will discuss the final method chosen, which was the possibility of using the spent foundry sand in asphalt production and the benefits to both companies in doing so.

This paper is a case study, which in part, forms research relating to the design of a model of continuous improvements in the cast metal industry.

1. INTRODUCTION

The UK Landfill tax was introduced by the Conservative Party in 1996. John Gummer was the Conservative Secretary of State and introduced the UK’s first environmental tax. Landfill tax is a method of enabling the United Kingdom to meet environmental goals, which were set out in the Landfill Directive that relates to the Land filling of biodegradable waste. By increasing landfill taxes, other waste treatment technologies with higher gate fees are more financially attractive to manufacturing companies. [1, 2]

In many countries, landfill tax is imposed on landfills and other disposal facilities as a means of raising revenue used for long term mitigation of environmental impacts relating to the disposal of waste. It is also used as a method of inhibiting disposal by raising the cost in comparison to preferable alternatives i.e. recycling. It is used in a similar manner to other government tax introductions such as excise or sin tax. [1-3].

Land filling is now discouraged for a number of reasons:

• Loss of recyclable materials, which could be reused but instead are dumped on landfills.
• Climate change, which is being accelerated due to gases produced at landfills from biodegradable waste and only adds to the green house effect.
• Constraints on areas suitable for landfill sites and a lack of suitable sites due to the ever increasing population, which makes land scarce.
• For companies it is becoming more and more expensive to dispose of waste due to the ever increasing landfill taxes.
The amount of tax levied is based on the weight of the material and whether the material is active or inactive. Inactive or inert waste covers materials such as concrete, brick, clay, soil, gravel, glass and with regards to this particular case, sand. Active waste covers all other material types such as wood, plastics and piping. Active and inactive waste is often mixed for example brick which can be mixed with wood. In these cases where contamination takes place the waste is usually treated as active. The standard rate of tax is levied on active waste and the lower rate is levied on the inactive waste. As a result of the categorisation of mixed waste as active this incurs the higher tax bracket. [2]

The landfill tax increase is putting pressure on manufacturing companies to find alternative methods of disposing of waste in an era where the environment and the economy are major global issues. In the United Kingdom 2008 budget, the landfill tax reached an all time high of £32 up from £24 the previous year. The 2007’s budget announced annual increases in the standard rate of landfill tax of £8 per tonne from 2008/2009 until at least 2010/2011, by which time it could be as much as £48 per tonne. The government announced in the budget that the aggregate levy, which has been frozen since its introduction, will rise in April 2008 from £1.60 to £1.95 per tonne.[2, 3]

The cost to a manufacturing company to dispose of waste is ever increasing. Therefore it was decided, in the local SME in question, that for a one year period the amount of waste taken to landfill would be monitored as a (KPI) Key Performance Indicator and new methods of recycling and reusing the waste would be investigated.

This case study was carried out in an SME that produces sand castings. There is a lot of spent foundry sand and general waste produced at the company. Due to the increase in landfill taxes scheduled until 2012, the company was forced to take some corrective action in disposing or recycling of the spent foundry sand. It was decided that this would be the basis of implementing the new Key Performance Indicators, which would relate directly to sand disposal. Over a period of one year the amount of sand and general waste that was accumulated per month was monitored along with the cost. During the monitoring process new methods of recycling or re-using the sand would be investigated and hopefully this would result in a reduction of the amount of sand being disposed of per month.

2. KEY PERFORMANCE INDICATORS INTRODUCTION

Key Performance Indicators or KPI’s which are also referred to as KSI’s, (Key Success Indicators) are created by many companies to help monitor and improve processes. The method of monitoring KPI’s is often referred to as BAM or Business Activity Monitoring. The method of monitoring and improving differs depending upon the organisation and the KPI in question but it is essential that they are quantifiable measurements that are agreed on prior to commencement, which will reflect on critical success factors for the company. They must reflect on the company’s goals and must be key to the company’s success. [9-11]

There are many different types of KPI used within businesses to measure keys goals such as increasing profit, increasing profit per employee, pre tax profit etc. With regards to the SME in question, the quantifiable key performance indicator would relate to the ever increasing cost of disposing of spent foundry sand. This was seen by the SME as Key to the future of the business. The cost of buying virgin foundry sand is increasing and so to is the cost of disposing of the spent foundry sand. This was pointed out in the government’s 2007 budget. Therefore the Key Performance Indicator that was generated aimed to reduce the percentage of sand taken to landfill. This was quantifiable (measurable) as every month the company received an invoice from the waste disposal company detailing the tonnage of sand disposed, the cost per tonne and the cost of delivering and collecting the skip loads of spent sand. [9-11]

However, measuring and monitoring the spent sand taken to landfill is singularly not the solution of reducing the percentage of sand taken to landfill. A new more environmental method of disposing or re-using the sand was required to achieve the companies overall goal of reducing the expenditure on sand disposal. The way in which the SME would measure the result would be to see how much sand month on month was reduced in tonnage sent to landfill and to also measure the overall cost of disposing of the sand in £. The KPI was implemented in September 2007 to reduce the amount of sand taken to landfill per month by 20% after one year, at which time a new goal would be reviewed or set. [9-11]
3. MEASURING AND MONITORING

Table 1.0 was derived from information received by the SME from the waste disposal company used to dispose of the spent foundry sand. It shows a period from April 2007 to April 2008. Information regarding sand tonnage disposed off, cost of disposal, container size, collection and haulage costs are all listed. As well as monitoring and measuring the sand disposed off, general waste and mixed contaminated waste was also measured. This was measured so that we monitor waste sand disposed which was mixed and therefore more expensive to dispose off and possibly examine the reasons behind it to find a solution.

4. KPI MONITORING AND MEASURING TABLE

Table 1.0: Detailed table of expenditure and tonnage involved in disposal

<table>
<thead>
<tr>
<th></th>
<th>Apr-07</th>
<th>May-07</th>
<th>Jun-07</th>
<th>Jul-07</th>
<th>Aug-07</th>
<th>Sep-07</th>
<th>Oct-07</th>
<th>Nov-07</th>
<th>Dec-07</th>
<th>Jan-08</th>
<th>Feb-08</th>
<th>Mar-08</th>
<th>Apr-08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sand Disposed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery of Containers</td>
<td>11.00</td>
<td>9.00</td>
<td>9.00</td>
<td>8.00</td>
<td>8.00</td>
<td>12.00</td>
<td>4.00</td>
<td>15.00</td>
<td>7.00</td>
<td>6.00</td>
<td>15.00</td>
<td>10.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Collection/ Haulage</td>
<td>11.00</td>
<td>9.00</td>
<td>9.00</td>
<td>8.00</td>
<td>8.00</td>
<td>12.00</td>
<td>5.00</td>
<td>15.00</td>
<td>7.00</td>
<td>6.00</td>
<td>15.00</td>
<td>10.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Disposal</td>
<td>11.00</td>
<td>9.00</td>
<td>9.00</td>
<td>8.00</td>
<td>8.00</td>
<td>12.00</td>
<td>4.00</td>
<td>15.00</td>
<td>7.00</td>
<td>6.00</td>
<td>15.00</td>
<td>10.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Sand tonnage taken away</td>
<td>148.20</td>
<td>128.68</td>
<td>128.26</td>
<td>109.36</td>
<td>123.40</td>
<td>172.68</td>
<td>56.78</td>
<td>201.40</td>
<td>95.52</td>
<td>79.52</td>
<td>208.90</td>
<td>139.58</td>
<td>204.98</td>
</tr>
<tr>
<td>Cost in £ for Delivery/Haulage</td>
<td>225.50</td>
<td>184.50</td>
<td>184.50</td>
<td>164.00</td>
<td>164.00</td>
<td>246.00</td>
<td>92.25</td>
<td>344.10</td>
<td>160.58</td>
<td>137.64</td>
<td>344.10</td>
<td>229.40</td>
<td>398.02</td>
</tr>
<tr>
<td>Cost in £ for Disposal of Waste</td>
<td>2,986.23</td>
<td>2,203.60</td>
<td>2,584.44</td>
<td>2,203.60</td>
<td>2,486.51</td>
<td>3,479.50</td>
<td>1,144.12</td>
<td>4,058.21</td>
<td>1,924.73</td>
<td>1,602.33</td>
<td>4,209.34</td>
<td>2,812.54</td>
<td>4,847.78</td>
</tr>
</tbody>
</table>

| General Waste Disposed |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Container Size CUM     | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   |
| Delivery of Containers | 1.00   | 0.00   | 1.00   | 0.00   | 1.00   | 0.00   | 0.00   | 3.00   | 0.00   | 0.00   | 1.00   | 0.00   | 1.00   |
| Collection/ Haulage    | 1.00   | 0.00   | 1.00   | 0.00   | 1.00   | 0.00   | 0.00   | 3.00   | 0.00   | 0.00   | 1.00   | 0.00   | 1.00   |
| Disposal               | 1.00   | 0.00   | 1.00   | 0.00   | 1.00   | 0.00   | 0.00   | 3.00   | 0.00   | 0.00   | 1.00   | 0.00   | 1.00   |
| General Waste Taken Away | 5.00   | 0.00   | 4.04   | 0.00   | 4.60   | 0.00   | 0.00   | 7.62   | 0.00   | 0.00   | 5.28   | 0.00   | 4.26   |
| Average                | 5.00   | 0.00   | 4.04   | 0.00   | 4.60   | 0.00   | 0.00   | 2.54   | 0.00   | 0.00   | 5.28   | 0.00   | 4.26   |
| Cost in £ for Delivery/Haulage | 69.53  | 69.53  | 69.53  | 69.53  | 69.53  | 0.00   | 0.00   | 212.25 | 0.00   | 0.00   | 70.75  | 0.00   | 77.26  |
| Cost in £ for Disposal of Waste | 205.75 | 166.25 | 189.29 | 0.00   | 313.56 | 0.00   | 0.00   | 217.27 | 0.00   | 224.29 |

| Mixed Contaminated Waste |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Container Size CUM     | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   | 6.10   |
| Delivery of Containers | 0.00   | 0.00   | 1.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Collection/ Haulage    | 0.00   | 0.00   | 1.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Disposal               | 0.00   | 0.00   | 1.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| General Waste Taken Away | 0.00   | 0.00   | 5.46   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Average                | 0.00   | 0.00   | 5.46   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Cost in £ for Delivery/Haulage | 0.00   | 0.00   | 39.50  | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Cost in £ for Disposal of Waste | 0.00   | 0.00   | 212.94 | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |

As mentioned, the sand reclamation KPI’s set up were used to monitor:

- The sand taken away per month
- The number of containers to be taken away
- The cost of disposing of the sand, which included the cost of delivering and collecting containers.
Flexible Automation and Intelligent Manufacturing, FAIM2009, Middlesbrough, England

Figure 1.0 graph shows the amount of sand taken away per month by the SME. The amount taken away is not consistent. The amount of containers taken away and also weight of the containers taken away is not consistent. In fact, one month shows an 18 tonne container was taken to landfill and another month shows that a container weighing only 11 tonnes was taken away. If each container could be filled to take away 18 tonne instead of 11 tonnes or the average 13 tonnes, then the number of container taken away could be reduced therefore reducing the cost to the company. This was one simple result to come from simply monitoring the sand taken away.

As well as the above mentioned topics, the main objective was to reduce the amount of sand taken to land fill by finding an alternative method which would be to find some means of recycling the spent foundry sand.

Figure 1.0: Cost in GBP OF Waste Disposal per Month

Figure 2.0 Cost in GBP of Delivery and Haulage

Figure 2.0 graph shows cost in GBP for disposing of waste sand from April 2007 to April 2008. The graph shows that the cost and therefore the sand amounts taken away over this period, are not consistent. By monitoring the sand using the KPI’s a number of key improvements could be made. If the tonnage of sand being taken away per month could be averaged out and become consistent, then the company’s finances with regards to outgoing expenditure could be predicted each month. At present the company cannot predict how much sand will be taken away in the next six months and how much this will cost.

In October 2007 the company spent £1,236.37 disposing of the waste sand, however a month later in the same year (before the price increase) the company spent £4,402.31. As a consequence of the company’s below average tonnage of disposed sand in October, they had no choice but to have more deliveries in November that inevitably cost the company more. The excess sand also took up far more space within the companies grounds.

Figure 3.0: Sand Tonnage taken to Landfill

Figure 4.0: Collections Made and Average Weight of Collection

Figure 3.0 and 4.0 graphs show sand tonnage taken away and the average weight of the collection per month. As an example, in March 2008, the company disposed of 200 tonnes of sand, however the average weight of collection was only 10 tonnes. This means the company had more excess sand within its workforce.
Monitoring the sand taken to landfill and monitoring the costs was also an advantage as price increases could be monitored. For example:

In April 2007 the cost of delivering and collecting one container was £41.00. In November 2007 it increased to £45.88 - an increase of 11.9%. This further increased in April 2008 to £56.86 - an increase of a further 23.93%. So from April 2007 to April 2008 the cost of delivery and collection of one container rose by 38.68%.

In April 2007 the cost of disposing of the sand was £20.15 per tonne that then rose in April 2008 to £28.43, an increase of £8.28 or 41%. The increases, according to the budget, are set annually until at least the year 2011 by which time it could reach £48 per tonne.

By monitoring the sand disposed of, it was also found that the one container load that was collected contained mixed waste. Mixed waste is categorised as general waste mixed in with the spent foundry sand. When this occurs the sand is treated as mixed waste and therefore levies the higher tax rate. Fortunately it was only 5 tonnes and not an average 13-14 tonnes. By Monitoring and creating the Key Performance Indicator a number of key improvements could be made to the SME’s current method of spent foundry sand disposal. Inevitably the SME wished to dispose of the sand for free via recycling, which would save the company a lot of money and also make the SME environmentally friendly.

Therefore it is important to not only find alternative methods of managing the waste disposed, but also to find alternative methods of reducing the spent foundry sand taken to landfill

4. A Investigation in Recycling Methods

There have been various case studies which show how reclaimed foundry sand can be used in various applications. The applications previously investigated by other researchers, SME’s and Large enterprises are listed in the table below.

Table 2.0: Uses of spent foundry sand [4-8, 12-13]

<table>
<thead>
<tr>
<th>Use</th>
<th>Notes</th>
<th>Type of Foundry Sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot Rolled Asphalt</td>
<td>Partial replacement in 50% mix of fine aggregates</td>
<td>Green Sand – Clay Content reduces bitumen bleed from mix</td>
</tr>
<tr>
<td></td>
<td>Well proven and successful application in UK and Overseas.</td>
<td>Alkaline Phenolic and resin shell sands are also suitable</td>
</tr>
<tr>
<td>Brick Manufacture</td>
<td>Used as an aggregate filler.</td>
<td>Most Sands including: Greensand.</td>
</tr>
<tr>
<td></td>
<td>Iron spotting on brick surface may cause continuity problems.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For standard Bricks – but may be a desirable for special effect bricks</td>
<td></td>
</tr>
<tr>
<td>Road base construction</td>
<td>Leaching of contaminants from unbound courses may pose problems – requires testing to ensure no adverse environmental impact.</td>
<td>Most chemically bonded sands may be used as substitutes for fine aggregates filler.</td>
</tr>
<tr>
<td>Concrete Block Making</td>
<td>Can be used in low density (aerated) and dense blocks.</td>
<td>Most sands including: Greensand.</td>
</tr>
<tr>
<td></td>
<td>Potential for phenol leaching from stockpile material may require modification of process authorisations</td>
<td></td>
</tr>
<tr>
<td>Foamed Concrete</td>
<td>Flowable fill, aerated concrete and controlled low strength material.</td>
<td>A percentage of greensand, alkaline phenolic and resin shell sands may be used as a fine aggregate replacement.</td>
</tr>
</tbody>
</table>
Table 2.0 briefly describes a number of tried and tested means of reclaiming spent foundry sand and a few methods that are still being investigated. The incentive of this research was to assess a number of potential methods of recycling the Alkaline Phenolic Sand produced by the SME in question. This would therefore allow attempts to recycle the spent foundry sand and reach the KPI goal of reducing the sand tonnage per month taken to landfill by 20% within 2 years.

The predominant uses that were examined were those which involved Alkaline Phenolic Sand. The company carried out chemical analysis on the sand prior to use and after use and found that a vast majority of the chemical binders were burnt off during the casting process. However it could not be determined how close the sand was to the liquid iron in the mould. Therefore it could not be certain whether the sand analysed had more or less chemicals burnt with it being closer or further away from the hot metal. It was decided that the method used to dispose of the sand would be an already tried and tested method of re-using spent foundry sand. This would allow the SME to meet its goal of reducing the sand tonnage taken to landfill faster. Further analysis of the sands chemical content was investigated along side the KPI.

The key uses above that use Alkaline Phenolic sand were in hot rolled asphalt, road based construction and foamed concrete. These areas were researched and relevant companies who manufacture such products were contacted.

5. THE PROCESS OF USING SAND IN ASPHALT PRODUCTION.

As part of the KPI investigation the SME made contact with a large enterprise which produces asphalt for the UK road system. It became apparent after a number of meetings that collaboration between the SME and the large enterprise would result in benefits for both parties.

The large enterprise was interested in using the spent foundry sand and substituting for virgin sand in the Asphalt production process. However before this could be achieved there were a number of tasks which had to be undertaken by both companies. The first task was to have the SME’s spent foundry sand analysed to see if it would be suitable for the standard currently beginning used by large enterprise.

The results that came back from the analysis laboratory were positive and proved that the SME’s foundry sand would be useable in the asphalt production process.

The second stage was to get permission from the local county council to use the sand. Once this was achieved a container load of spent foundry sand was taken from the SME by the large enterprise to be screened and crushed. This enabled the large enterprise to examine the content of the spent sand to ensure foreign objects were not passed though and added to the sand in the asphalt mix. The grain size of the sand was also important to the large enterprise as this can determine the major properties of the asphalt. The large Enterprise carried out tests using various percentages of virgin and spent foundry sand to determine the correct percentage of spent foundry sand that would be used in the asphalt production.

<table>
<thead>
<tr>
<th>Sand Bags</th>
<th>Spent Greensand, potentially other sand types pending investigations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foundry sand has the potential to be used in sandbags in emergencies. Potential for phenol leaching requires further investigations.</td>
<td></td>
</tr>
<tr>
<td>Cement</td>
<td>A percentage of greensand, alkaline phenolic and resin shell sands may be used as a fine aggregate replacement.</td>
</tr>
<tr>
<td>Cement uses sand and potentially spent foundry sand can be used to produce it. However, as with foamed concrete, the strength of the cement can decrease because of residual particles of resin that may adhere to the sand grains.</td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>Potentially Greensand.</td>
</tr>
<tr>
<td>Certain foundry sand types can be mixed with various soils for the benefit of agriculture. Further investigation and tests are required.</td>
<td></td>
</tr>
</tbody>
</table>
The introduction of a KPI to monitor Spent Foundry Sand disposal

The sand price was then negotiated between the two companies so that both companies would benefit from the exercise. A price of £4 per tonne was agreed to be paid by the SME. This would cover the cost of collecting the spent foundry sand by the large enterprise and transporting the sand to the asphalt production plant.

6. Results

Table 3.0: The annual cost to dispose of spent foundry sand from April 2007 to March 2008

<table>
<thead>
<tr>
<th>Year</th>
<th>Waste Type</th>
<th>Price Per Tonne</th>
<th>Collect and Delivery Of Skip</th>
<th>Skip Size Cu. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Sand</td>
<td>20.15</td>
<td>20.50</td>
<td>12.20</td>
</tr>
<tr>
<td>2008</td>
<td>Sand</td>
<td>23.65</td>
<td>25.49</td>
<td>12.20</td>
</tr>
</tbody>
</table>

April 07-March 08 | 132.69 | Average Sand in Tonnes Disposed Per Month
April 07-March 08 | 115    | Loads taken per year
April 07-March 08 | 1,592.28 | Tonnage sent to Landfill

April 07-March 08 | £32,084.44 | Total annual cost per tonne
April 07-March 08 | £2812.54  | Annual cost of collection and delivery
Total             | £34,561.01 | Total Cost (Excluding VAT)

| Large Enterprise | New Price | £4.00 | Per Tonne |
| Annual Cost based on April 07-March 08 figures. | £6369.12 | Annual Cost Quote |
| Saving when compared to April 07-March 08 Total Costs | £28,191.89 |

Figure 5.0. Shows Monitoring of the KPI after the implementation of the Sand Reclamation.
8. DISCUSSION

The average cost to dispose of the sand per month from April 2007 to 2008 was £2,840.94. This is represented on figure 5 with a red line. The goal of the KPI was to reduce the sand taken to landfill per month by 20% within 2 years represented on the graph by the green line. Since the large enterprise started recycling the spent sand the cost of disposing of the spent sand has reduced to an average of £580 per month. Not only has this venture between the two companies saved the SME money but it has also saved the large enterprise money too. Now the large enterprise no longer has to purchase as much virgin sand to produce the asphalt. A case study provided by the company showed that in 1997 a similar venture saved the company £31,500 by not paying for virgin sand to be delivered to the company. [14]

The information in figure 5.0 shows how much annually it cost the company to dispose of the sand prior to the implementation of a Key Performance Indicators. The total cost of disposing of the sand in 2007 was approximately £34,561. If the SME uses the same amount of sand every year and the cost per tonne remains at £4 the SME can expect to pay £6369.12. This equates to a saving of £28,191.89 year on year.

The KPI drove the company to not only successfully achieve but greatly exceed the yearly goal of a reduction in sand disposed off at landfill per month by 20% by year one. As of August 2008, 100% of the sand taken away per month is recycled and re-used in other products and the company saves an average of 79.5% per month compared to average amount of sand taken away between April 2007 and March 2008.

9. CONCLUSION

The implementation of a Key Performance Indicator at an SME proved to be successful. The objective of reducing the sand taken to landfill by 20% in year one was not only completed but exceeded. The SME now recycles 100% of its spent foundry sand which was previously taken to landfill.

There are a number of other benefits to come from KPI implementation. The SME has now established a working relationship with a large enterprise. It can now continue to monitor the KPI and try to ensure that the sand taken away per month is averaged out and therefore costing the same each month.

The large enterprise will benefit from reusing the sand saving money on purchasing virgin sand for production purposes.

The research has also been environmentally beneficial. The SME is no longer dumping spent Alkaline Phenolic sand in the landfill thus prolonging the life of the landfill by not allowing it to become saturated with sand, which helps to reduce the amount of landfill sites created throughout the country. The trucks used to transport the skip...
loads of sand used to travel to the company leave an empty skip and travel back, and then travel to collect the full skip and travel back. The new company sends a truck collects the sand and travels back only once, reducing the amount of trips required and therefore saving on CO2 emissions.

REFERENCES:


[8] Rafat Siddique and Albert Noumowe Department of Civil Engineering, University of Cergy Pontoise, 5, Mail Gay Lussac, Neuville sur Oise, 95031 Cergy-Pontoise, France Utilization of spent foundry sand in controlled low-strength materials and concrete


[13] Environmental Best Practice Programme, Sand Transfer Eliminates Virgin Sand Use, A case study from Joseph and Jesse Siddons Ltd

[14] Aggregate The Sustainable aggregate information service, Cost Benefits Case Study from Tarmac UK Ltd.
Modelling and Implementing Circular Saw Blade Stone Cutting Processes in STEP-NC

Julio Garrido Campos* and Ricardo Marín Martín
Automation and Systems Engineering Department
University of Vigo
Vigo, 36200, Spain

ABSTRACT

The current paper describes a STEP-NC compliant implementation of circular saw blade stone cutting machining processes. Although some stone machining processes have already been covered in the STEP-NC research and standardization initiatives (as for instance stone machining through stone milling machines), there have not been yet, however, any detailed model proposal to cover circular saw blade stone cutting operations. There are some major differences between conventional metal cutting and stone cutting using a circular saw blade, as well as some important technology parameters. The paper highlights these main differences as well as some other relevant characteristics of the circular saw blade stone cutting machining operations, and proposes a STEP-NC extended model to take them into account. The paper presents an implementation for these machining operations with a STEP-NC compliant saw blade stone cutting machine prototype. A definition of the product features is performed with a prototype CAM system, and it is communicated to the controller together with the rest of machining data (based on the proposed model) in a STEP-21 file. The machine control system processes the machining data (after converting the part 21 file into a STEP part 28 file) and drives the machining of the features by implementing the operations defined in the model.

1. INTRODUCTION

For stone construction parts, such as balusters, columns, mouldings etc (figures number 1, 2, 3 in Figure 1), circular sawblades are the most popular cutting mechanism [1]. Sawblade stone cutting machines are very similar, despite the fact that they do vary in terms of mechanical configuration and in terms of control [2]. The cutting process is a general sequence of activities based on a fairly standard approach applied worldwide and performed in three main phases (Figure 2). In the first phase, cuts are made progressively with cut depth following the outline or profile of the final part (the desired surface) and leaving a specific distance between cuts. The sawblade makes a complete longitudinal cut and goes downward step by step, making several parallel cuts at different depths (passes). Once it reaches the desired depth, the sawblade disc exits the stone and moves to start another parallel cut at a given distance from the previous cut. Second phase of eliminating the stone between the cuts is typically a manual operation, implemented using tools as mallets, chisels, etc. Once the material between cuts (slides) is removed, the third process phase obtains the final surface, where the rough and terraced profile obtained from the previous phase is smoothed by making overlapping cuts.

Processing stone is a multidimensional and complex task in which, factors such as physical material properties, sawblade characteristics, sawing conditions, cooling efficiency, etc., are all interrelated and influence process efficiency and quality. Different conditions during the cutting process (tool cutting power variations, changes in the stone structure, etc., [3]) is a relevant aspect of this technology, so, and from the machine automation point of view, to get an optimized process, the control or/and the operator should have the ability of making deep changes in the middle of the process [4]. Other relevant aspect of the technology is that using discs limit the number of axis which can be moved when they are into the stone (only moves in the cutting plane can be done), although to make

* Corresponding author: Tel.: (0034) 986-812610; Fax: (0034) 986-814014; E-mail: jgarri@uvigo.es
complex features as arc patterns (Figure 1), even five axis movements may be needed and performed when the disc is out of the stone.

Disc or circular saw stone cutting machines may be, therefore, very complex, with 5 or more axis (Figure 3) to be able to machine that complex parts and, in some cases, are being programmed using the same technology as their metal CNC machines counterparts, the G&M codes [5].

A wide range of CNC manufacturing processes have already been addressed by research initiatives and projects with a view to extending the STEP-NC standard [6]. There have also been research initiatives with regard to stone machining. Stroud [7] has addressed stone milling operations and has defined new architectural stone features. Part 15 of the ISO 14649 standard—although aimed primarily at wood and glass processing—defined some stone sawblade operations, although the approach is very general. Therefore, no detailed research into STEP-NC and circular sawblade stone cutting operations has as yet been reported. Any sawblade stone cutting machine implementation related with STEP-NC has been reported either.

This paper proposes a STEP-NC compliant CAM/CNC architecture to automate disc stone cutting machines. After a brief introduction to STEP-NC technology parts (section 2 of the paper), a extended model is introduced in section 3 in order to get a STEP-NC compliant model for saw blade cutting processes. This extended model is based on the selection and definition of new features and on the modelling of the machining process. The resulting model is the base for the development of the STEP-NC stone cutting CAM and CNC machine (section 4), based on the definition of features to be communicated to the controller and machining operations parameters to make them. The architecture is designed to be able to react to changes in the machining conditions, very common in this technology.
The controller has the objective of machining the features, and it is able to re-planning, on real time, the work to get them despite changing conditions in the stone or in the disc.

2. STEP-NC TECHNOLOGY PARTS

Today, a new standard often known as STEP-NC is being developed to provide a data model for a new breed of intelligent CNC Controllers [6].

Two versions of STEP-NC are being developed. The first is the Application Reference Model (ARM), as ISO 14649 [8]. ISO 14649 models are written in EXPRESS language [9], while designs are stored in ISO 10303-21 physical file format [10]. The second STEP-NC version, the ISO 10303 AP-238 [11], has adopted the ARM models built in ISO 14649 as the ARMs, and it is defined the AIM (Application Reference Model) to build a common language with other STEP data models.

ISO 14649 represents the STEP-NC ARM, with the definition of the detailed information requirements that are to be fulfilled. Part 1 is an overview and fundamental principles definition, while Part 10 [12] is for general process data to provide a set of basic capabilities for process planning for machined parts. Besides these two general parts, other parts are dedicated to specific technologies: Part 11 [13] is the process data for milling, Part 12 [14] is the process data for turning. Additional parts define the specific technology tools, as Part 111 for milling tools, Part 121 [15] for turning tools, etc. (Figure 4).

In order to expand the use of STEP-NC new technology are taken into account in new parts as, for instance, to EDM (Electro Discharge Machining) [16] [17] by defining a complete extended model of STEP-NC that became ISO 14649-13; to closed loop manufacturing processes for inspection that became a new ISO 14649-16; ISO 14649-15 to define process data for contour cutting of wood and glass, etc. [18]. Other technologies just use the current state of the standard [19], as for instance, the dry high-speed milling of marble and industrial ceramic (the LITHO-PRO project), in which the STEP-NC milling model has been used, but defining a new range of process-specific features [7]. Sawblade cutting technology for stone parts have several specific parameters with no clear equivalent in milling, turning, etc. technologies. To explicitly consider them in the STEP-NC model, new technology entities would have to be added (a new ARM definition from an ISO 14649 perspective for the sawblade stone cutting technology). A new prototype model is presented next.

3. STEP-NC FOR SAWBLADE STONE CUTTING

The new ARM data model proposal is built upon the basic STEP-NC process model (ISO 14649-10) and in accordance with Part 11 for milling and Part 12 for turning. The model defines technology-specific data types representing stone cutting processes following the basic STEP-NC approach of separating geometric and technological information. The extended model uses features already defined in the standard for milling and turning and in other research projects [20]. In accordance with other technological parts (for example, Part 11 for milling) and with new developments (for example, Part 16 for inspection), the model defines so-called sawblade cutting working steps, including machining operations and features to be machined. The main model entities for the proposal are described in Figure 4.

STEP-NC programs can be described as a set of complex and structured tasks (workingsteps) to machine the features of a workpiece [21]. Figure 4 represents a STEP-NC program organised around a main workplan containing a series of working steps. Each workingstep applies a specific machining operation to manufacturing features using specific tools, sets of technology parameters and specific strategies.

Features specify the information necessary to identify shapes of interest in a mechanical product. These shapes represent volumes of material either removed by machining operations or resulting from a series of machining operations (ISO 14649-1). The sawblade stone cutting basic feature is the cut-out, which may be described with the slot feature already defined in ISO 14649-10 and also in ISO 10303 AP-224 [22], as well as the ISO 14649-12 cut-in, defined as a slot. Two main kinds of complex features are considered: planar features for mouldings and turning features for columns and balusters.
Figure 4: Sawblade stone cutting ARM general overview (ISO 14649 style)

Figure 5 represents also a simplified model for sawblade stone cutting operations, and main machining strategies for some of the operations. Surface operations are used to machine features for a fixed stock and also to machine planar faces in an indexed feature. When performing a sawblade_turning_operation, the stock has to rotate, either continuously to obtain revolving features through revolving_turning_operations, or step-by-step to obtain indexed features through indexing_turning_operations. In both cases—and as happens with the surface features—roughing and finishing operations are executed.

Different strategies may be adopted in each operation. Figure 4 represents some of the strategies associated with the sawblade_surface_operation, while other strategies (not illustrated) can be defined for turning operations.

As with the standard for milling and turning, the model is completed with other information definitions—technology_parameters, approach_retract_strategies, machine_functions, etc.—. For example, a technology entity would include technology parameters as feed rate value, sawblade speed value, Boolean values to allow or disallow feedrate override, etc. Finally, information would also be needed on the main tool parameters, for example, sawblade diameter, maximum allowed blade depth in the stone, etc.

4. IMPLEMENTATION

A first STEP-NC prototype CAM-CNC system was implemented with the extended model described above and based on a commercial 5 axis circular sawblade stone cutting machine. Contrary to the current NC programming standard, STEP-NC is not a method for part programming and does not normally describe the tool movements for a CNC machine. Instead, it provides an object oriented data model for CNCs with a detailed and structured data interface that incorporates feature-based programming [23]. The prototype STEP-NC control architecture follows the CAM-embedded STEP-NC controller structure defined by Zhang [24] where the controller is a combination of a STEP-NC code interpreter, a basic CAM and a NC controller (the interpreter translates the physical file into internal data format and the CAM makes decisions on machine-specific details and generates low-level control commands that are executed by the NC control). Zhang architecture is represented in Figure 5 (central part).
The original machining system architecture (Figure 6) is composed of a proprietary CAD/CAM-embedded system for selecting feature and machine parameters, although profiles for features may be imported from CAD files in DXF format. The CAM information is translated to XML (eXtensible Markup Language) and communicated to the HMI (Human Machine Interface or MMI: Man Machine Interface). Both CAD/CAM and the HMI run in the same embedded PC under Windows CE, but the first could well be run in another computer as the HMI system can also import the XML files. The HMI adds shop-floor information to the XML file (for example, current sawblade diameter). The resulting file is the input for the machine low-level control module, responsible for axis motion control, alarm management and input/output management. This module runs in a TwinCAT PLC Run-Time system with TwinCAT NC axis control and is programmed in IEC 1131.

Although the system looks like a classical two stages architecture [25] (the first is the CAM programming stage; the second is the actual machine stage carried out in a CNC machine tool), much of the intelligent is on the CNC side. This means that the high level (Windows CE planning process) does not communicate the low level (the CNC controller) the trajectories to make a feature, but it communicates the geometric data of them. It is the CNC system itself who translates that “high level program” to a “low level program” where the tool movements sequence is specified. The controller, depending on the feature parameters (geometry), decides the sequence of cuts to perform (the tool path). Figure 5 right represents the controller structure, following the Zhang approach.

In the prototype (Figure 6), a CAM application allows features and operation parameters to be selected and generates a STEP Part 21 file following the AP-238 extended model. This file is transformed to a STEP Part 28 format [26] by a STEPPart21-to-STEPPart28 translator and is communicated to the machine. Next is an example of the translation of a STEP-part 21 – program line and its equivalent following the STEP Part 28 specification.

![Figure 5: Control architecture](image)

![Figure 6: Implementation architecture](image)
The machine-embedded CAM/HMI system in Figure 6 first processes the XML Part 28 file—taking advantage of XML file access libraries—and a post-processor translates the data into the machine information system, matching feature operation parameters coming from the STEP-NC file to the feature operation parameters of the machine.

4.1. AN EXAMPLE

Figure 7 represents an example where a change in the machining conditions occurs. Theses changes are very frequent in the stone cutting technology, mainly because of the non-homogeneous nature of the stone and also because of the diamonds saw blades dynamic behavior. So, it may occur that the process has planned a specific tool path trajectory (with a specific feed rate, a specific step length to go downward in the stone, and a specific cut distance between two parallel cuts) depending on parameters as the stone hardness. But if that hardness changes (increases) in some area of the stone or if the disc suddenly loses its cutting efficiency, it would make harder for the disc to cut the stone. The machine may be able to automatically realize of the new conditions through several parameters (motor current, real feed speed from the encoders...step 5 in Figure 5). To be able to go on with the cut, it would be necessary to change the disc speed (easy to do), but may be also needed to recalculate the tool path, for instance, with a narrower distance between the cuts (in Z, in Y or both).

This can be made by the control system, dynamically, because it knows the geometry of the feature being machined. It has the algorithms that, given the feature as an input, plus other parameters coming from the machine sensors, and of course, knowing the amount of the work already done (the current position), generates new tool paths to feed the motion control task (step 3 in Figure 5).

Figure 7 is an example for the main sequential steps to machine a part: from the design to the machining, which is performed by the machine system of Figure 6 with the control structure of Figure 5. Figure 7 (a) represents the feature geometric information coming from a CAD system (activity “1” in Figure 5 performed by the Figure 6 CAD/CAM proprietary system): in this case a linear molding. This information is communicated to the machine control system through a STEP XML file. This file is postprocessed (activity 2 in Figure 5), and with the information the Machine Control high level algorithms, plans the machining work to be performed (activity 3 in Figure 5): it decide tool paths, tool speeds and feed rates (Figure 7 b). This information is used for the Machine Control low level algorithms to perform the final run-time motion control (activity 4 in Figure 5). The example also represents a change in the cutting conditions in the middle of the machining process: for instance, a change in the disc cutting conditions (Figure 7 "c")

In the example, the machine operator realizes of the new situation and, based on his experience, decides not only slow down the speed rate of the process, but also to change the number of vertical cuts, as well as he decides to reduce the distance between two cuts. As result, the planning algorithm recalculates and generates a new “program” for the remaining of the work (step “d” in Figure 7).

5. CONCLUSIONS

As the example has illustrated, the control system has to have all the information needed in order to support on-line dynamic process planning. This information is composed of CAM data (geometric information about features
and machining parameters), and real-time machining process conditions. While the second is directly acceded by the controller through its sensors, the first has to be communicated from the design phase. Finally, as conclusion, algorithms to plan the work (tool path calculation) and algorithms to manage the work (motion order sequencing and dispatcher) should all rely in the low level to be dynamic and allow on-process work. Also the algorithms, additionally to the feature geometric data, they should have a feedback input from the machine to continuously drive the process of calculating tool paths to make the part (the feature).

The architecture presented in the paper is the first step to be able to have a STEP-NC based stone disc cutting machine. To do that, a review of already defined in STEP-NC features and features proposed in other STEP-NC stone related works and new defined features, will be needed to transform the one the machine is able to take into account. Also, new model entities should be defined to take into account specific technology parameter, as for instance, the cutting process parameters. As result of the process, a new disc cutting STEP-NC compliant extension would be proposed.

The aim of the paper is not to provide a definitive work on STEP-NC sawblade stone cutting, but rather to present one possible approach. To explicitly consider them into the STEP-NC standard, ISO 14649 style ARM model would need to be approved by the ISO SC1 committee to become a new ISO 14649-## part, and ISO SC4 committee would have to expand the AP-238 ARM, AIM and mapping tables. However, much more work would be needed to be done before in order to cover the complete process chain from design to process planning and manufacturing. To reach this point, all CNC automated processes in a stone processing plant would need to fit in STEP-NC. Also, more feedback from other stone cutting systems implementers will need to be considered.

It could be said that a Conformance Class 3 (CC3) approach has been used to make the prototype STEP-NC disc cutting machine as the CAM communicates the controller not the tool path but the feature geometry. ISO 10303 provides for a number of options that may be supported by an implementation grouped into Conformance Classes: CC1 for tool path programming; CC2 for closed-loop programming; CC3 for feature based programming and CC4 for generative programming. CC1 supports the description of machining programs containing a single sequence of operations, each of which is described using the machine-independent path of the tool centre point, using a simplified set of curves types, as well as tool requirements, management information about the program, and technology-specific parameters. Conformance Class 2, which expands the previous one to support the description of machining programs with full range of toolpath specifications as well as full shape information for the workpiece, rawpiece and restricted areas on the setup. Conformance Class 3 extends the previous one to support the description of machining programs using a full range of executable constructs and manufacturing process features defined by implicit parameters. With this Conformance Class, a CAM would just provide the feature, and that should be enough for the CNC to machine the part. The prototype controller machine architecture described in previous section works similar to CC3, but without being conformed with CC1, although that limitation would not be a hard implementation problem.

ACKNOWLEDGEMENTS

This work is supported by the MCYT (Ministerio de Ciencia Y Tecnología) as project DPI2006-05772.

REFERENCES


Ink-Jet Printing System for Manufacturing Conductivity Patterns

Hee-Kwan Lee¹, Dong-Won Kim²*, Gyun-Eui Yang³, Eun-Young Heo², Jong-Yeong Lee²

¹Automobile Parts & Mold Technology Innovation Center, Palbokdong2ga, Duckjingu, Jeonju, 561-855, South Korea

²Department of Industrial & Information Systems Engineering and The Research Center of Industrial Technology, Chonbuk National University, Jeonju, 561-756, South Korea

³Department of Mechanical Engineering, Chonbuk National University, Jeonju, 561-756, South Korea

ABSTRACT

Ink-jet based printing methods, using conductivity molecules or nano-particles, are receiving much attention thanks to the appearance of electronic papers, organic electroluminescent (EL) displays, and flexible printed circuit boards (PCB) in the commercial market. These leading technologies and products require new processes with know-how which can produce circuit patterns through simple and reliable methods under low cost. This study introduces an ink-jet based printing system (IJPS) for manufacturing functional PCB patterns, which takes advantage of a multi-nozzle ink-jet head discharging conductivity polymers or nano-particles. The IJPS is composed of a chemical materials supply line, a printing head control device, an XY precision-movement stage, and patterned circuits. Consequently, it can print an area of 1,200 mm x 1,300 mm with a resolution of 600 x 1,600 dpi and at printing speeds of 6 m² per hour. A printing head can use either a 128 or a 256 nozzle head. Through test experiments, conductivity inks have been printed onto polyamide films, PET films, and glass plates, and then the conductivities of the printed patterns have been verified.

1. INTRODUCTION

Since early 2000, ink-jet based printing methods using conductivity molecules or nano-particles have been the focus of much research with several leading industries changing to the commercial goods stage from the R&D stage, with items such as, electronic papers, organic electroluminescent (EL) displays, solar batteries, and flexible printed circuit boards (PCBs). The ink-jet printing technology (IJPT) is more price competitive than the existing semiconductor manufacturing processes relying on photolithography, and chemical or plasma etching [1-2], due to its economic use of functional materials that facilitates mass and low-cost production [3]. IJPT offers powerful capabilities such as printing narrow lines and surface-bound gradients, and can even carry a programmable feature to change the sequence of operations to accommodate different product configurations [4].

An industrial ink-jet system is attractive to the market, which can print functional patterns by discharging conductivity polymers or nano-particles through a set of printing nozzles. In this respect a multi-nozzle printing head is especially useful from this sense. Moreover, the ink-jet heads can be applied to various key industries, such as, direct PCB manufacturing lines for the part materials industry, OLEDs (Organic Light Emitting Diode) lines for the displays industry, industrial bio-sensor manufacturing lines using a constant amount ink discharging device, printers for the advertising industry, and flat panel displays.

All these application fields require complex electric and electronic circuits on PCBs. From this aspect, using the growing know-how allows for the production of circuit patterns, through easy and reliable methods which are very important since the knowledge can be used to activate and extend the future-trends in the industry market with high values. OLEDs and bio-sensor based flexible PCBs in particular use various materials such as silicon, glass fibers, and polymers whose manufacturing processes are relatively complicated. Typically, a circuit pattern manufacturing process in these fields is composed as follows: metal thin film coating → photo resist formation → photo masking (film) → UV...
light scanning → development → chemical etching → film retrieval → post-processing → wiring board. These steps can be simplified considerably by adopting IJPT to the following: board with various materials → pre-processing → ink-jet pattern formation → post-processing → wiring board. This study aims to develop an ink-jet based printing system that can discharge conductivity materials that is able to support the part materials sectors, leading displays, and bio-sensor manufacturing industries.

2. **State of Art Technologies**

An ink-jet printing system (IJPS) has both mechanical and electrical parts. The former is composed of a movement stage, and a printing head with nozzles, while the latter has a motion control and ink-shooting unit. The electrostatic parameters of the electrical part include a voltage and its offset, and frequency and duty cycles. Teng and Karim [5] experimented on the voltage and nozzle cone size for optimal ink jetting, considering other parameters such as the ink viscosity, applied voltage, nozzle diameter, and applied pressure. Rahman *et al.* [6] addressed the droplet generation accuracy when ink fluid was being jetted from the printing system. Several researchers focused on fine patterning to apply color filters, LEDs, and circuit boards into display instruments and semiconductors [7-9]. However, further study on different printing system arrangements and strategies is needed to satisfy the requirements of ink materials and printing mechanisms.

This study aims to develop a low-resolution, large-scale IJPS for application to cinemas with big screens that are rapidly increasing in the market, by taking into account the printing speed of IJPS and the precision movement of a gantry-type stage.

**Figure 1:** Pattern manufacturing by conductivity materials ink-jet (Source: Spectra)

Figure 1 shows IJPT including its conductivity materials and application areas. Philips and Markem have investigated circuit pattern manufacturing, including PCB and LCD, by discharging conductivity materials. Litrex developed a commercial, OLED manufacturing technology by discharging organic and conductivity materials, adopting a piezo-ink-jet printing head.

Conventionally, OLEDs are manufactured by either a low-polymer or a high-polymer. Low-polymer OLEDs are manufactured through existing, vacuum-metallization, while the high-polymer OLEDs are manufactured by an ink-jet method. Figure 2 shows an OLED pattern manufacturing using the ink-jet method, which is regarded, by many researchers and companies, to be the top leading technology globally [8].

**Figure 2:** OLED manufacturing by ink-jet methods (Source: Spectra)
3. **IJPS for PCB Pattern Manufacturing**

This study proposes IJPS that prints PCB patterns, as shown in Figure 3. The IJPS is composed of a chemical material supply line, a head control device, an XY-precision movement stage, and a circuit board for producing a particular pattern.

![Figure 3: Composition of the IJPS](image)

### 3.1. **Ink-Jet Head Control**

A printing head, made by Spectra’s SL series, has been used in this study. The input voltage can be from 0 to 150V DC and the power supply of 120V is used for the head (See Figure 4). Figure 5 shows the change of ink-jet amount according to vibrational frequencies. The board for driving the head can control the allowed voltage and its wave pattern and frequencies, since these values determine the ink-jet amounts and the line widths of the patterns.

![Figure 4: Ink-jet head for discharging conductivity materials](image)

![Figure 5: Relationships of ink-jet amount with discharging speed according to allowed frequency](image)
3.2. Chemical Materials Supply Line

The ink is supplied to an ink-jet head by a pump, and is transported through a tube that is free from chemical change or other contaminations; the path of the ink’s movement is as follows: main ink bottle → pump → filter → ink line → auxiliary ink bottle → printing head (see Figure 6). Of course, the ink is not affected by the ink pump, bottles or filters which have been selected by considering ink ingredients and the characteristics required.

Figure 6: Supply of conductivity ink to the printing head

3.3. XY-Precision Movement Stage

For the stage lineup, a linear motion is chosen by considering a linear motor, and a screw or a belt type. The XY-precision stage is composed of two or three sets of linear motions. The linear motion control has to support a resolution of 1 μm by using linear scales and sensors. Then, the precision motion control is performed by the feedback from the linear sensors. A flat-type linear motion is manufactured and tested for the printing of the patterns. The XY-precision stage adopts a gantry type structure, so as to minimize the positioning error caused by the deformation and inertia from the self-weight, which is occurred in acceleration and deceleration.

Figure 7: XY-stage

3.4. Printing Head Control through a Computer

An interface is required between a computer and an ink-jet printing head so as to process and control digital signals, as shown in Figure 8. A PCI9054 chip is used for the interface, and printing data is sequentially transported through a FPGA (field programmable gate array) device.
4. A Case Example

A prototype IJPS has been developed and tested to verify the performance of the system.

![Interface between a computer and a printing head](image1)

**Figure 8: Interface between a computer and a printing head**

4.1 XY-Stage

The XY-stage is developed on a pallet of 1,200 mm x 1,300 mm so that a screw or belt type linear motion can be installed (see Figure 9). The pallet uses an optical plate with a flatness of 100 µm and the printing head can scan the X-axis first, and then take feed motion along the Y-axis, using as small an increment as the output resolution can support. Conventionally, the movement resolution of the X-axis requires around 600 to 800 dpi (dots per inch), and an X-axis scanning speed of about 2 m/sec.

![Drawing of the XY-stage](image2)

**Figure 9: Drawing of the XY-stage**

4.2 Printing Head Control Device

A printing head control device consists of a carriage module, including an XY-stage with a printing head, a control box module controlling the entire system and transmitting data, and a hydraulic and pneumatic module supplying ink with constant pressure.

X and Y-axes motions are made by a belt and a screw type motion, respectively. The X-axis has been manufactured by extrusion and can produce X-axis motion by installing a linear motion guide on an upper and a side wall. An AC servo motor is used for the axis-driving, and a linear encoder, with a resolution of 150 dpi, enables the motor to secure a resolution of 600 dpi. Likewise, a screw type Y-axis linear motion is driven by an AC servo that supports 1,600 dpi.

A carriage is formed by installing a carriage block on a jig that is mounted on an XY-stage base (see Figure 10 and 11). The carriage block is composed of a base plate, ink storage, a circuit fixing board, and a linear motion block moving along a linear guide. The base plate is formed by a head and a head-fixing jig, having a structure that controls...
the vertical and horizontal planes of the head. A printing nozzle is made by a set of two pieces of Spectra’s S class. Nova 256 is enough for single-ink nozzles. The carriage base has 6 partitions carrying a maximum of 6 types of ink. The carriage has many different parts such as mechanical, electronic, and piping in a small space, requiring delicate design and precision assembly of the parts.

Figure 10: XY-motion on a plate stage

Figure 11: Carriage with a head and a circuit board

The controller section is divided into 3 sub-parts, namely, an electric apparatus with a power supply, a central, electronic part that controls the printing system, and a motion drive that dictates linear motions. A power supply supports 5, ±15, 24, and 120 volts of DC power sources. The 5 volt power supply is used for the logic chips and 24 volts for the valves, cartridges and heaters. A digital to analog converter uses a power supply of ±15 volts with an operational amplifier, in order to generate constant waveforms into the head. The 120 volt DC power supply is used for driving a piezo-electric chip that supplies constant electric waveforms into the head. A motion drive part works in conformity with signals from an electronic control system, displaying information about the moving positions. The motion drive part consists of an AC power source and control signal lights wired to signals of a motor encoder.

Finally, a prototype IJPS is completed, as shown in Figure 12. The system can print a pattern of 1,200 mm x 1,300 mm, supporting the resolution of 600 x 1,600 dpi, printing the pattern at 6 m² per hour. A printing head can take either 128 or 256 nozzles, with a drop size of 80 μm and a printing speed of about 10 kHz.

Figure 12: Prototype IJPS
4.3 EXPERIMENTAL RESULTS

Commercial conductivity Ag nano-ink, made by Cabot Co., has been prepared for testing the prototype IJPS. The physical and chemical characteristics of the ink are shown in Table 1. Poly-amid films, popularly used for flexible PCBs, PET films, and glass plates, are available for the base materials for IJPS.

Table 1: Characteristics of Ag nano-ink

<table>
<thead>
<tr>
<th>Items</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viscosity</td>
<td>14.4 cP at 22°C</td>
</tr>
<tr>
<td>Surface tension</td>
<td>31 dynes/cm at 25°C</td>
</tr>
<tr>
<td>Solid loading</td>
<td>20 wt %</td>
</tr>
<tr>
<td>Density</td>
<td>1.24 g/cc</td>
</tr>
<tr>
<td>Particle diameter</td>
<td>30 - 50 nm</td>
</tr>
<tr>
<td>Vehicle</td>
<td>ethylene glycol based</td>
</tr>
</tbody>
</table>

Figure 13 shows the following screen shots for a patterning experiment: (a) an interface program for the design of an image and its delivery to a printing head, (b) a printing motion, (c) a printed image, (d) and (e) scanned images to verify the performance of the developed system, and (f) pattern measurement. From the image measurement, the width of the printed pattern (d) ranged from 124 to 162 μm, which satisfies the tolerance of ±20μm from a base width of 145 μm. The heights of the base and of the pattern base are about 10 nm and 60nm, respectively, while the deviation (e) of the height that forms the pattern base is also within ±20μm, which satisfies the requirement for a big display.

(a) printing interface (b) printing motion (c) printed pattern

(d) pattern width (e) pattern height (f) pattern measurement

Figure 13: Patterning experiment
5. SUMMARY

An IJPS has been developed in this study for manufacturing functional PCB patterns. The system takes advantage of a multi-nozzle ink-jet head discharging conductivity polymers or nano-particles. The system is composed of a chemical material supply line, a printing head control system, an XY precision-transferring stage, and pattern circuits. It can print an area of 1,200 mm x 1,300 mm with a resolution of 600 x 1,600 dpi and printing speeds of 6 m² per hour. A printing head can take either a 128- or 256-nozzle head. In the test experiments, conductivity inks were printed by using the developed system onto a polyamide film, a PET film, and a glass plate. The conductivity of the printed patterns was verified with sufficient precision tolerances for application to big screen displays.

ACKNOWLEDGEMENT

This study was financially supported by the research grant (IR2006-2) of Chonbuk National University, Jeonju, South Korea.

REFERENCES

Determine the Optimum Settings for Maximum Force ($F_{\text{max}}$) on Quad-Flat-Pack’s (QFP’s) using a Wetting Balance Machine.

Robert Woods*, Christy Gillick

Department of Manufacturing & Operation Engineering, University of Limerick, Limerick, Ireland

ABSTRACT

Solderability is an extremely important factor for electronics assembly. The majority of solder defects are attributable to poor (insufficient) component solderability. Testing for solderability is necessary. Within many electronics manufacturing companies a lot of focus is aimed at reducing the amount of soldering defects by carrying out sample solderability testing on batches of components using the Wetting Balance test method.

Feedback from a variety of component suppliers indicate the main factors that influence the soldering characteristics for the Wetting Balance test are: Solder Temperature, Immersion Depth, Immersion Speed, and Dwell Time. Similar to the Dip and Look solderability test, there are various International Standards with varying settings for the aforementioned factors when trying to evaluate a components soldering ability using a Wetting Balance machine. A review of the International Standards such as J-STD-002C, IEC60068-2-54 and IEC68-2-69 was required to obtain the range of settings used for each of the factors. A matrix of all the settings within the standards was developed and from this a Design of Experiments (DOE) was set-up and conducted. It was decided to investigate the Wetting Balance test in more detail paying particular attention to Quad-Flat-Packs (QFP’s). This paper focuses on the $F_{\text{max}}$, maximum force reached during the Wetting Balance test. $F_{\text{max}}$ is one of a number of responses associated with the Wetting Balance test. Others include $T_b$ (time to buoyancy), $T_{F_{\text{max}}}$ (time to reach $F_{\text{max}}$), $F_1$ (force after 2 seconds) and $F_2$ (force after 5 seconds). The effect each of the four variables has on $F_{\text{max}}$ is determined by Main Effects and Interaction Plots.

DOE 1 was set-up using Minitab and a total of 48 runs were carried out for the four factors. Each factor had two levels, high and low. A Pareto Chart of the Standardised effects showed all four factors had an effect on the $F_{\text{max}}$. Immersion Speed had the largest effect - as the speed increases from 1mm/sec to 5mm/sec the $F_{\text{max}}$ increases. DOE 2 provided a more in-depth analysis of the four factors. Each factor had three levels, low, medium and high resulting in 81 runs. Again analysis was carried out using Minitab. The investigations and analysis from DOE 2 determined the optimum settings for a high $F_{\text{max}}$ using a Wetting Balance machine. These settings are Solder Temperature - 230°C, Immersion Speed – 0.5mm/sec, Immersion Depth – 0.5mm and Dwell Time – 4 seconds.

1. INTRODUCTION

The solderability of a component’s metallic termination is a critical parameter in any soldering process because it represents the likelihood of that termination forming a good alloy with the solder and a high quality solder fillet. As components continue to become smaller and more fine pitch, coupled with the widespread use of less active fluxes, the soldering process window narrows and the impact of poor solderability increases. Although components are generally assembled from parts of known (good) solderability, there is no way of guaranteeing this without testing a sample batch using the Wetting Balance Machine to access the components soldering characteristics [1]. The Wetting Balance Machine is a purpose designed instrument for testing components such as QFP’s, SMT and also pads of Printed-Circuit-Boards. In essence, a Wetting Balance Machine exploits the fact that if a metallic body is dipped into a bath of molten solder, the force and speed with which the solder meniscus climbs upwards on the body’s immersed surface indicates how well the solder wets it and thus its solderability. In simple terms, the greater

*Robert Woods: E-mail: BobbyWoods@topmail.ie
the solderability, the higher a meniscus will climb, which can be measured as a change in the vertical force acting on the suspended specimen, in this case a component lead. [2]

The Dip and Look test is another quick method of testing the solderability of a component but previous research and investigation has deemed this test to be invalid. [3] Only a pass/fail result is given by the Dip and Look test whereas the Wetting Balance investigates the components soldering characteristics.

For the purpose of this investigation the maximum force (Fmax) reached during the Wetting Balance test will be focused on. The Fmax is measured in milli-newtons (mN) by the Wetting Balance machine. Fmax is the maximum force reached during the Wetting Balance test, i.e. the maximum force required for the solder to wet to the lead of the component. [2] A high Fmax result on the Wetting Balance Machine is a high attraction of force required for solder to attach to the component lead. The objective of this paper is to develop the optimum settings to achieve the highest Fmax allowable with the range of settings within the International Standards.

2. REVIEW OF INTERNATIONAL STANDARDS

The main International Standards reviewed for this paper are:
- J-STD-002C – Solderability Test for Component Leads, Terminations, Lugs, Terminals and Wires.
- IEC60068-2-54 – Test Ta: Solderability testing of electronic components by the Wetting Balance method.
- IEC68-2-69 – Test Te: Solderability testing of electronic components for surface mount technology by the Wetting Balance method.

Table 1 is a matrix of the settings got from the aforementioned standards. A high and low value was used for each factor.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Min Value</th>
<th>Max Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solder temperature (°C)</td>
<td>2</td>
<td>250</td>
</tr>
<tr>
<td>Immersion speed (mm/sec)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Immersion depth (mm)</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>Dwell time (sec)</td>
<td>4</td>
<td>10</td>
</tr>
</tbody>
</table>

The immersion speeds stated in IEC68-2-69 (1mm/sec to 5 mm/sec) was taken due to the fact that the speed specified in J-STD-002C and IEC60068-2-54 (1mm/sec to 20mm/sec) was too fast for the MUST II machine at the high end (20mm/sec). This resulted in the machine going into an ‘error fault’ failing to record any data.

3. DESIGN OF EXPERIMENT (DOE)

There are various Wetting Balance test parameter settings available for solderability testing using the Wetting Balance method; however there is no clear consensus as to the appropriate combination of parameters which define an acceptable Fmax standard at component level. In order to achieve a workable standard, a review of each available International Standard is required to develop a Design of Experiments with a view to obtaining the optimum settings. The goal of this experiment is to determine the critical variables and any interactions present during the Wetting Balance testing of components. The resulting settings from the DOE should provide an optimum set of parameter settings to achieve the highest possible Fmax in order to detect any form of contamination on the termination of the components which could affect solderability.

The software used to conduct this D.O.E is ‘Minitab V13.

The components investigated were Quad Flat Pack’s (QFP) using the MUST II Wetting Balance Machine. No preconditioning of the components was carried out prior to the testing and all components used in the test were from the same supplier batch and date code.

*Robert Woods: E-mail: BobbyWoods@topmail.ie
Determine the Optimum Settings for Maximum Force (Fmax) on Quad-Flat-Pack’s (QFP’s) using a Wetting Balance Machine

3.1 RESPONSES

The maximum force (Fmax) will be analysed by Design of Experiments. Fmax is a very important response to determine a component’s ability to solder. During the Wetting Balance test, the machine automatically records the maximum force reached and this is graphically represented by the machine. Three replicates were conducted for each array.

3.2 FACTORS AND RUN COMBINATIONS

Minitab developed the array of factors and settings used in the experiment. These factors were determined by reviewing the international standards as stated in section 2. As stated earlier, the main factors of the Wetting Balance test are:

i. Solder Temperature – temperature at which the solder of the Wetting Balance machine is set at. SnPb alloy used

ii. Immersion Depth – the depth the component is immersed into the SnPb solder

iii. Immersion Speed – the speed at which the component enters the solder

iv. Dwell Time – amount of the time the component is immersed in the solder

These four factors had two levels, high and low (see table 1), resulting in 48 runs.

At the beginning of every run, the machine settings were made with reference to the arrays determined by Minitab. Once steady state conditions were achieved and settings verified, the run was initiated. Each run consisted of three replicates. One lead of a QFP was used for each replicate resulting in three replicates for each array. After the test was completed, each result for Fmax was automatically recorded by the Wetting Balance machine. For each array the results were inserted into Minitab in column “Fmax”.

3.3 ANALYSIS OF DESIGN OF EXPERIMENTS

Figure 1 displays a Pareto of standardised effects for Fmax. It can be seen that all four factors show some form of effect on the Fmax. The magnitudes of the effects are represented by the position on the Pareto, i.e. the interaction of BD (Immersion Speed and Dwell Time) has the largest influence on the Fmax but B (Immersion Speed) on its own has little or no effect. All interactions to the right of the red line will be investigated further using Main Effect (Figure 2) and an Interaction Plots (Figure 3).

Figure 1 Pareto Chart of the Standardised Effects for F\textsubscript{max}
Determine the Optimum Settings for Maximum Force (Fmax) on Quad-Flat-Pack’s (QFP’s) using a Wetting Balance Machine

The main effects that the four factors have on the Fmax are shown in Figure 2.

- Immersion Speed has the biggest effect. As the speed increases from 1mm/sec to 5mm/sec the Fmax increases. The range is approximately 0.1mN.
- Solder Temperature decreases the Fmax as it is increased from 230°C to 250°C with a range of 0.035mN approximately.
- As the Immersion Depth is increased from 0.25mm to 0.5mm the Fmax increases with a range of 0.03mN approximately.
- Varying the Dwell time from 4 to 10 seconds has little effect on the Fmax.

The interactions of Solder Temperature and Dwell Time, Immersion Speed and Dwell Time, and Immersion Depth and Dwell Time have been established as those that have a significant effect on the Fmax. Figure 2 displays the two way interaction of all four variables. The interactions with parallel line indicate that the interaction does not affect the Fmax.

*Robert Woods: E-mail: BobbyWoods@topmail.ie
3.4 CONCLUSIONS FROM DESIGN OF EXPERIMENTS 1

By focusing on the interactions that are significant the following is evident:

- Solder Temperature and Dwell Time – A constant solder temperature of 230°C and increasing the dwell time from 4 to 10 seconds will result in a decrease in Fmax. A constant solder temperature of 250°C and varying the dwell time from 4 to 10 seconds increases the Fmax result.

- Immersion Speed and Dwell Time – By maintaining an immersion speed of 5 mm/sec and varying the dwell time from 4 to 10 seconds cause the Fmax to decrease but by maintaining an immersion speed of 1 mm/sec and varying the dwell time from 4 to 10 seconds cause the Fmax to increase.

- Immersion Depth and Dwell Time – By maintaining an immersion depth of 0.5 mm and varying the dwell time from 4 to 10 seconds cause the Fmax to decrease but by maintaining an immersion depth of 0.25 mm and varying the dwell from 4 to 10 seconds cause the Fmax to increase.

The next step in the experimentation involved planning a second DOE to provide a deeper understanding of the Maximum Force (Fmax) reached during the Wetting Balance test.

3.5 DESIGN OF EXPERIMENTS 2

DOE 2 was planned using Minitab and the objective was to generate a set of parameters on a Wetting Balance machine that would give a low Fmax of the Wetting Balance test. The four variables from DOE 1 were selected and experimented at 3 levels, ‘Low’, ‘Medium’ and ‘High’. All other variables were constant, i.e. solder, component type and all other Wetting Balance machine parameters. 81 runs were made with one replicate. The reason for only using one replicate for DOE 2 was due to the fact that no variation was seen for DOE 1 and Gauge R&R study carried out on the machine prior to the experiment was within the limit of 10% [2]. Table 2 shows the low, medium and high settings that were use.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solder Temperature</td>
<td>230</td>
<td>240</td>
<td>250</td>
</tr>
<tr>
<td>Immersion Speed</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Immersion Depth</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>Dwell Time</td>
<td>4</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

Figure 4 and 5 show the Main effects and Interaction plots as a result of DOE 2 for Fmax. The main effects the four variables have on the Fmax are shown in Figure 4.
Similar to DOE 1, Immersion Speed has the biggest affect on the four variables. Immersion Depth shows little or no affect.

- **Solder Temperature** – As the temperature is increased the Fmax decreases, with a range of $0.21$ mN approximately.
- **Immersion Speed** – As the speed is increased the Fmax decreases, with a range of $0.34$ mN approximately.
- **Immersion Depth** – Very little effect on Fmax.
- **Dwell time** – the effect of dwell time is very small similar to that of the immersion depth. As dwell time increases from 4 to 10 seconds the Fmax decrease but increasing from 10 to 15 the Fmax increases.

Figure 5 Interaction Plots for DOE2

### 3.6 Conclusions from Design of Experiments 2

Again focusing on the interactions that were significant from DOE 1, the following conclusions can be made;

- **Solder Temperature and Dwell Time** – By maintaining a constant temperature of $230^\circ$C and varying the dwell time from 4 to 10 seconds there is very little affect on Fmax but the Fmax results are highest at this temperature. A dwell time of 10 to 15 seconds cause the Fmax to decrease. A constant temperature of $240^\circ$C and a dwell time variation of 4 to 10 seconds cause Fmax to increase but a dwell time from 10 to 15 seconds result in an increase of Fmax. A temperature of $250^\circ$C has little of no affect on Fmax with a dwell time variation of 4 to 15 seconds and gives the lowest Fmax results.

- **Immersion Speed and Dwell Time** – A constant immersion speed of 0.1mm/sec and a varying dwell time from 4 to 10 seconds show a slight increase in Fmax but a dwell time from 10 to 15 seconds decrease the Fmax. An immersion speed of 0.3mm/sec with dwell time variation from 4-15 seconds has minimal affect on Fmax. A high Fmax result is evident with an immersion speed of 0.5 mm/sec. A reduction in Fmax is seen from 4-10 seconds and the opposite effect is seen from 10-15 seconds. Overall an immersion speed of 0.5mm/sec gives the highest Fmax results.

- **Immersion Depth and Dwell Time** – It can be concluded that the interaction between immersion depth and dwell time is not as significant from the results in the Main Effect graph in Figure 4 and also the Pareto of the Standardised Effects in Figure 1. For the purpose of this investigation it can be stated that an immersion depth of 0.5mm and a dwell time of 4 seconds will give the highest Fmax result. This can be seen from the Main Effects graph in Figure 4.

*Robert Woods: E-mail: BobbyWoods@topmail.ie*
3.7 SUMMARY OF FACTOR SETTINGS FOR HIGH FMAX

Table 3 is a summary of the settings for each of the factors to achieve the optimum high Fmax value using a Wetting Balance machine.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solder Temperature</td>
<td>230°C</td>
</tr>
<tr>
<td>Immersion Speed</td>
<td>0.5 mm/sec</td>
</tr>
<tr>
<td>Immersion Depth</td>
<td>0.5 mm</td>
</tr>
<tr>
<td>Dwell Time</td>
<td>4 seconds</td>
</tr>
</tbody>
</table>

Fmax provides important information in determining the soldering characteristics of a QFP lead in a tin/lead (Sn/Pb) process. A high maximum force (Fmax) result got by using the Wetting Balance Machine represents a high attraction of force required for solder to attach to the component lead. The higher the Fmax the more difficult the solder has to attach to the lead of the component under test.

3.8 REFERENCES

Understanding the Dynamic Response of a Machine Tool Structure

by Finite Element Analysis

Q. Xu,¹ and S. Barrans²

¹School of Science & Technology
University of Teesside
Middlesbrough, Tees Valley, TS1 3BA, UK

²School of Computing and Engineering
University of Huddersfield
Huddersfield, HD1 3DA, UK

Abstract

This paper presents research work on the structural dynamic behaviour of an existing CNC machine tool through finite element analysis (FEA). This type of knowledge is needed for achieving high machining accuracy and high productivity. Modelling techniques are developed and parametric study is carried out. The predicted dynamic response from FEA is in good agreement with the results from experimental mode testing. This research also proves the reliability of finite element analysis which can be used for the optimal design of new machine tool structure.

1. Introduction

The competitive nature of the global economy demands industry to maximize the efficiency and quality, as well as minimise response time and costs in production [1-2].

It is understood that the machining dynamic process involves both the machine tool structure and the work-piece [2]. However, this research focuses on the machine tool structure part.

A machine tool system has three main groups of parts: mechanical structures, drives, and controllers. All three groups affect the relative position of the cutting tool and work-piece. This paper concentrates on the mechanical structure where deformation at the contact point may be due to thermal effects, the mass of the work-piece and machine tool components, and cutting loads. A substantial amount of work has been carried out into the static and pseudo-static response of machines to such loads. However, relatively little work has been carried out on the dynamic response. This paper will report research work on the structural dynamic behavior of an existing CNC machine tool. The general references for finite element analysis are ample and can be found in such as [3-4].

The development of an appropriate finite element (FE) model and the selection of nominal values for the stiffness of various components (such as ball-screws and bearings) are reported. The techniques used to model the various structural elements are also discussed. This model has been validated against modal test data [5].

The paper then describes a series of parametric studies of the dynamic behavior of the machine tool, concentrating principally on the mode shapes and natural frequencies. The parameters considered are: 1) ball-screw stiffness; 2) bearing stiffness; 3) slide head positions; 4) dimensions of the structural components; and 5) the material properties of the structural components. Through this parametric study, the overall effects of those parameters on the mode shapes and natural frequencies were obtained. Carrying out this type of study by experiment is impractical. Careful analysis of these results can provide guidance on how to improve the machine’s performance.

The final part of the paper describes the steady state response of the machine under an assumed excitation force. The application of modal damping, validated by modal testing, is described. The characteristics of the results obtained are in good agreement with real cutting experiments and offer an explanation of the occurrence of significant errors at a frequency lower than the first natural frequency.
2. Master Model Development

A master FE model was developed first where the selection of nominal values for the stiffness of various components (such as ball-screws and bearings) is reported. Various modelling techniques applicable for dynamic analysis were employed. Following careful analysis of the structure of the specific machine tool, it was decided to use 4 different types of element as shown in Table 1. The total number of element was 2911 and the number of node was 13367. The mesh is shown in Figure 1.

A typical value of ball-screw stiffness is 194.8 kN/mm calculated according to the cross sectional area and the typical value of Young’s Modulus. The nominal values for bearing stiffness on the x and y-axes and on the column were 500 KN/mm, 420 KN/mm and 530 KN/mm, respectively. The head was positioned at about the mid-span of the column in the z direction. The table was positioned in the middle of the x motion and close to the column in the y direction, as shown in Figure 1.

The ABAQUS finite element software [6] was used for the analysis and FemGV [7] was used for pre- and post-processing. Modal analysis was conducted on the specific FE model and the first 10 mode shapes and natural frequencies were obtained. The main feature of the first mode shape is the nodding of the head in the y-z plane. The value of the first natural frequency, 25.4 HZ, was in good agreement with 28 HZ from experimental modal testing [5]. It was concluded that the FE model developed was an accurate model and suitable for dynamic analysis. The mode shape is not presented here for brevity.

<table>
<thead>
<tr>
<th>Plates</th>
<th>QU8-S8r5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodies</td>
<td>He20-C3d20</td>
</tr>
<tr>
<td>Bearings</td>
<td>BE2-springa</td>
</tr>
<tr>
<td>Ball Screws</td>
<td>Be2-springa</td>
</tr>
<tr>
<td>Plates</td>
<td>QU8-S8r5</td>
</tr>
<tr>
<td>Bodies</td>
<td>He20-C3d20</td>
</tr>
<tr>
<td>Bearings</td>
<td>BE2-springa</td>
</tr>
<tr>
<td>Ball Screws</td>
<td>Be2-springa</td>
</tr>
</tbody>
</table>

Table 1 List of types of element used

Figure 1: The mesh generated for the machine tool
3. PARAMETRIC STUDY

The first group of parameters were stiffness of the bearings and ball screws. These were chosen as:

Ball Screw Stiffness: 300, 250, 194.8, 150 and 10 KN/mm;
The Bearing Stiffness on Column: 420 to 606 KN/mm;
The Bearing Stiffness on X-axis: 490 to 570 KN/mm;
The Bearing Stiffness on Y-axis: 380 to 500 KN/mm.

The typical results for the effects of the bearing stiffness and ball screw stiffness are shown in Figures 2 to 3.

The second parameter was the head position. Analyses were carried out with the head positioned at the mid-span of the column and furthest from the table. The results obtained are shown in Figure 4.

Figure 2: The effect of bearing stiffness on the natural frequencies

Figure 3: The effect of ball screw stiffness on the natural frequencies
The third parameter varied was the wall thickness of the column. Values of 6 mm, 9 mm, and 12 mm were used. The results are shown in Table 2.

### Table 2 The effect of thickness of column on the natural frequencies

<table>
<thead>
<tr>
<th>Wall Thickness</th>
<th>Mode 1</th>
<th>Mode 2</th>
<th>Mode 3</th>
<th>Mode 4</th>
<th>Mode 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mm</td>
<td>25.4 HZ</td>
<td>25.9 HZ</td>
<td>30.4 HZ</td>
<td>62.3 HZ</td>
<td>77.6 HZ</td>
</tr>
<tr>
<td>9 mm</td>
<td>30.5 HZ</td>
<td>39.8 HZ</td>
<td>45.2 HZ</td>
<td>67.2 HZ</td>
<td>83.5 HZ</td>
</tr>
<tr>
<td>12 mm</td>
<td>34.5 HZ</td>
<td>50.8 HZ</td>
<td>60.1 HZ</td>
<td>70.5 HZ</td>
<td>86.4 HZ</td>
</tr>
<tr>
<td>Change %</td>
<td>36 %</td>
<td>96 %</td>
<td>98 %</td>
<td>13 %</td>
<td>11 %</td>
</tr>
</tbody>
</table>

The fourth parametric study was to examine the effect of filling the column and the base with concrete. Here it was assumed that the concrete and the wall adhered perfectly and no rebounding occurs. Filling the column with concrete is labelled as Case 1, filling the base only is designated Case 2 and filling both is identified as Case 3. The obtained natural frequencies are given in Table 4.

### Table 3 The effect of filling concrete on the natural frequencies

<table>
<thead>
<tr>
<th>Mode No.</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>master</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>27.1</td>
<td>25.4</td>
<td>31.8</td>
<td>25.4</td>
</tr>
<tr>
<td>8</td>
<td>35.8</td>
<td>29.9</td>
<td>39.8</td>
<td>25.9</td>
</tr>
<tr>
<td>9</td>
<td>41.3</td>
<td>30.3</td>
<td>41.7</td>
<td>30.4</td>
</tr>
<tr>
<td>10</td>
<td>47.3</td>
<td>77.1</td>
<td>66.5</td>
<td>62.3</td>
</tr>
<tr>
<td>11</td>
<td>78.6</td>
<td>80.2</td>
<td>81.7</td>
<td>77.6</td>
</tr>
<tr>
<td>12</td>
<td>80.7</td>
<td>82.4</td>
<td>83.9</td>
<td>79.8</td>
</tr>
<tr>
<td>13</td>
<td>82.9</td>
<td>84.3</td>
<td>85.4</td>
<td>81.1</td>
</tr>
<tr>
<td>14</td>
<td>84.3</td>
<td>85.5</td>
<td>88.5</td>
<td>84.3</td>
</tr>
</tbody>
</table>

### 4. STEADY STATE RESPONSE

The steady state response was investigated. Forces were applied at the tool tip in the x and y directions and were defined as having a sinusoidal variation with a changing frequency. A constant value for modal damping ratio was
used in this analysis. The displacement response spectrums for the x, y, and z directions were obtained and are shown in Figure 5.

From Figure 5 it is clear that the largest displacement in x direction, y direction and z direction occurs when the input force frequency is 18 HZ, 20 HZ, and 20 HZ, respectively.

It would be interesting to repeat the FE simulation using experimentally obtained modal damping ratios. However, the authors are aware of the problems in determining modal damping ratios by modal testing.

A real cutting experiment, conducted by colleagues, revealed that the largest displacement occurred at lower frequencies. This FE analysis may be used to explain why this happened. Further digestion of these FE results will enable design modifications to improve the machine’s structural dynamic stiffness.

5. CONCLUSIONS

Finite element based dynamic analysis of a machine tool structure has been conducted. A master FE model was developed and validated first. Then a series of parametric studies were conducted to examine the effects of a number of variables such as bearing and ball screw stiffness, head positions, wall thickness, and filling the column and base with concrete. Further investigation was conducted into the steady state response.

This paper contributes to 1) the general methodology of using FEA in the machine tool design process; 2) knowledge of the overall machine dynamic behaviour under various configuration; and 3) specific information on how to improve machine design.

Figure 5: Response to sinusoidal force input at tool tip.
REFERENCES


Methodology of soldering process defects management

Frantisek Steiner *
Department of Technologies and Measurement
University of West Bohemia
Pilsen, CZ30614, Czech Republic

ABSTRACT

Paper deals with defect problems of soldering process. One of the most important processes of electronic equipment production is the soldering process. The reason for this statement is the fact that the prospective dependability of electronic equipments depends above all on soldered joint quality. Poor quality of soldered joint can caused equipment failure in the future. That is why we will pay the attention to these problems. Paper will present the proposal of methodology itself and its application. This proposal comes from the procedure of process dependability improvement. Many factors are concerned in resulting soldered joint quality. These factors can be divided into six basic categories: material, information, procedures, machines and accessories, human factor, and process environment. The methodology includes the way of risk analysis as well as the way of identified risk management. Further, an incident management is the part of proposed solution. This way describes activities necessary to incident treatment, i.e. identification of defect causes, their eliminations and the minimization of possible defect impact. Finally, the structure of knowledge base, which is generated by data collection in the frame of incident management, is designed. Records about the occurrence, history and the way of incident solving, collected in the knowledge base, help us to solve new occurrence of the same or similar incidents. The purpose of the knowledge base formation is also objectification of risk analysis process inputs.

1. INTRODUCTION

Soldering process is one of key processes in electronic equipment production. Final equipment quality and mainly dependability is influenced above all by this technology. Soldering process, which could seem to be simple, is influenced by number of factors, which affect resulting soldered joint. That is the cause that soldering process is one of the most exacting processes.

Thanks to miniaturizing spacing and package size the soldering process is almost on the imaginary limits of physical possibilities. But that is not sole problem, which is object of next development and research. In recent years ecological aspects assert oneself in production. Implementation of device contacting technology without toxic lead is consequence of this. Either classical lead based solder is replaced by lead-free alloys or conductive and non-conductive adhesives are used. The adhesives are used not so often due their prices. Therefore we can mainly encounter lead-free soldering processes in practice.

Replacement of classical lead based solder by its alternative presents a number of troubles and problems. Lead-free alloy properties approach to lead based solder properties only. It is mostly necessary to use alloys of several elements and the lead-free solder choice often presents compromise, because the alloy choice with one better property mostly means worse second one.

In context of above-mentioned problems concerning solder processes with the inclusion of lead-free technology implementation, raising of integration and possible failure consequence in mass production it is necessary to prevent defects or at the least to identify them quickly after formation and to remove their causes as soon as possible. In context of mass production, the losses related to imperfect soldered joints present far more financial losses. Soldered joints are often impossible to repair due to above-mentioned higher integration. One of the reasons is lower dependability of repaired soldered joint but main reason is impossibility of hand rework due to very small size of device packages. There is possible to create quality soldered joints by machine soldering at first attempt. Any hand

* Corresponding author: Tel.: (+420) 377 634535; Fax: (+420) 377 634502; E-mail: steiner@ket.zcu.cz
rework or second attempt of machine soldering is not good. The price of printed circuit boards and the price of mounted devices must be also included in contingent financial loss.

Following text will describe the methodology proposal. Proposed methodology appears from process dependability approach. This methodology unlike other in common used methods takes into account the importance of information for soldering process management. The methodology tries to move an attention to the area of defect prevention or very fast reaction on occurred defects. Risk management system was suggested. This system will analyze possible defects and than will suggest proper action. Eventually occurred defects will be analyzed by incident management and the causes of their formation will be determined. The information obtained from this analysis will be stored for next usage. We can use it for analysis of next occurrence of same or similar failures.

2. PROCESS DEPENDABILITY

2.1. PROCESS DEPENDABILITY UNDERSTANDING

Process dependability is different from technical systems dependability. Technical systems dependability is only one of possible impacts effecting process dependability. Process dependability analysis extends the enterprise process management concept according standards series ISO 9000:2000. Process dependability care is one of continuous process improvement aspects in enterprise.

Process dependability is more comprehensive conception than technical systems dependability. There are shown symptoms of the system “man – machine – environment” in the field process dependability. Parts “man” and “environment” are understood in the field of technical systems dependability rather marginally as process conditions, or requirements for servicing and maintenance. For example a failure caused by the service is not comprehended like equipment undependability but like outside factor. On the other hand the process dependability conception includes all of these three parts. On the basis of generally true experiences it is supposed, that human factor will have dominant influence on process dependability, at least at specific process types. [2], [3]

Process dependability importance grows up also due to present-day trends. Consumer constraint to price cutting and the emphasis on accuracy of delivery are these trends. Late delivery as well as early delivery is for consumer undesirable. Late delivery can cause production interrupt and by contrast early delivery requires sufficient consumer warehouse capacity.

A business process optimization in term of process dependability is solving of these problems. Than we will have sureness (though with specific amount of residual risk) that business inputs will be transformed to outputs (products) in planned time and consumer will be also satisfied in planned time.

The dependability is process quality sign as well as product quality sign. Process dependability at the same time reflects by specific way the process stability. Therefore the process dependability assessment should become the most important element for process capability assessment. [2], [3]

2.2. INCREASES IN PROCESS DEPENDABILITY

There is possible (according [2]) to divide the impacts on the process dependability into six basic categories, which are related to 6 basic questions (Fig. 1):

1. **Material**: What is it made of?
2. **Information**: Which information is it based on?
3. **Procedures**: How is it made?
4. **Machines and accessories**: What is it made with?
5. **Human factor**: Who is made?
6. **Process environment**: Where is it made?

There is suitable to divide the measures for process dependability increasing also according these categories.
Specialized publications show following steps for increasing of total enterprise process dependability [2]:

1. Total enterprise process decomposition on sub-processes.
2. Categorization of sub-processes according to their function:
   - Main processes,
   - Support processes,
   - Control processes.
3. Sequence and order analysis of mentioned processes in term of:
   - Inputs and outputs of processes,
   - Time interval (sequence, parallel and serial processes).
4. Determination of critical processes in term of:
   - Significance (e.g. FMEA methodology),
   - Time (e.g. CPM - Critical Path Method),
   - Substitutability.
5. Critical processes decomposition on particular parts (e.g. Ishikawa diagram) – Fig. 1:

| a) input material quality, |
| b) input information dependability and quality, |
| c) production procedures and documentation quality, |
| d) dependability of machines, equipments and furniture, |
| e) human factor dependability, |
| f) environment quality. |

6. Determination of the factors which are critical in specified sub-processes.
7. Choice of suitable method for possibility analysis of critical factors dependability increasing.
8. Plan and realization of improvement.
9. The efficiency check.

3. PROPOSED METHODOLOGY

Framework of proposed methodology is described in Table 1. Choice of suitable analysis type is first activity of this methodology. Several possibilities are offered here. In principle it is decision if material factor will be analyzed by the help of elementary analysis or by the help of detailed risk analysis. In the case, that material analysis will be made by elementary method, activities of quality input material factor analysis follow.

Application of selected detailed risk analysis method is next step of methodology. Risk catalogue and risk treatment plan are risk analysis results. Incident management is last part of proposed methodology. The knowledge base will be created within the frame of incident management. This knowledge base can be used in several ways.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Output</th>
<th>Methods, tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice of analysis type</td>
<td>Risk analysis procedure description</td>
<td>Soldering process assessment in manufacturing context</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Orientation analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elementary analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-formal analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Detailed risk analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined analysis</td>
</tr>
<tr>
<td>Elementary analysis of input material quality factor</td>
<td>Material quality assessment including contingent avoidance of poor quality material using</td>
<td>Solderability testing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wetting balance method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Globule method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rotary dip method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Meniscus rise method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Area of spread method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Edge dip method</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dip method</td>
</tr>
<tr>
<td>Detailed risk analysis making by the help of selected method</td>
<td>Risk catalogue</td>
<td>Methods of detailed risk analysis</td>
</tr>
<tr>
<td></td>
<td>Risk treatment plan</td>
<td></td>
</tr>
<tr>
<td>Incident management implementation in soldering process</td>
<td>Incident records in knowledge base</td>
<td>Data collection on the basis incident management</td>
</tr>
</tbody>
</table>

Diagram of proposed methodology is also shown in Figure 2. How it is evident from figure, the methodology can be divided in two parts, risk management and incident management. First part is focused on risk management of defect event. The task of this part is identification of possible risks and proposal of measures. These measures will eliminate non acceptable risks. Risk management can be divided in two parts, risk analysis and risk treatment. Risk catalogue and risk treatment plan are outputs of risk management. Risk catalogue contains all identified risks and we have to decide if certain risk can be accepted or not. After non acceptable risks determination the risk treatment plan can be designed. The plan contains all measures for risk elimination. It is possible to say that risk management part demarcates the area of defect event prevention in soldering processes.
Second part, i.e. incident management (Fig. 2), represents enlargement of classical frame of dependability analysis. Incident management defines procedure how incident will be solved and also how we will learn from it. The imaginary output of incident management part is individual incident occurrence list. This list represents real incident origination probability. The real incident probability means more objective background for next risk analysis. By this way it is stepwise possible to objectify relatively subjective risk analysis results. The great numbers of incidents have to take effect of risk increasing and subsequent measure. On the contrary it is also possible to check applied measures effectiveness.

The significant benefit of incident management part is also knowledge base formation. Knowledge base will be used for next incident solution. That means we can use information saved in knowledge base for faster response to new incident occurrence. Data from knowledge base help us to identify incident causes and thereby we can correct them faster.

4. APPLICATION OF METHODOLOGY

4.1 RISK MANAGEMENT

HVA method (Hazard Vulnerability Analysis) was used for risk analysis. Example of this method is shown in Table 2. At first the assets, threats and vulnerabilities are identified. In terms of identified treats and vulnerabilities the incident probability (PI) is derived. Then the impact of incident (D) on certain asset is assessed. For incident probability and impact assessment we can use an empirical scale, e.g. 1 – 10. Risks are calculated as $R = PI \times D$. The last column named Measures is used for recording of present measures. Already established measures affect incident probability and impact of course. The advantage of this method is well-arranged form including summary of present measures.

4.2 INCIDENT MANAGEMENT

In previous part risk analysis was presented. But it is necessary to provide the data for determination and assessment of risks. It is necessary to know incident probability and impact of incident. These parameters can be estimated of course. But the estimation is not accurate.

The best way, how to obtain data for risk analysis, is incident management application. Incident management gives us data for risk analysis but also it help us to get incident under control by easier way (faster and with smaller impacts).

Within the frame of incident management it is possible to establish knowledge base, which will contain so-called “known errors”. Records of information about incident beginning, continuance and solving are used for next incident solving. The structure of knowledge base is shown in Table 3.
Table 2 Hazard Vulnerability Analysis

<table>
<thead>
<tr>
<th>Asset type</th>
<th>Assets</th>
<th>Threats</th>
<th>Vulnerabilities</th>
<th>Incident probability (PI)</th>
<th>Impact (D)</th>
<th>Risk (R)</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solder paste</td>
<td>5</td>
<td>Change of asset property</td>
<td>Non-performance of storage conditions</td>
<td>9</td>
<td>5</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low quality material</td>
<td>Insufficient input control</td>
<td>5</td>
<td>5</td>
<td>25</td>
<td>Paste from prestigious producer</td>
</tr>
<tr>
<td>PCB</td>
<td>5</td>
<td>Change of asset property</td>
<td>Non-performance of storage conditions</td>
<td>6</td>
<td>5</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low quality material</td>
<td>Insufficient input control</td>
<td>10</td>
<td>5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Oven thermal profile</td>
<td>4</td>
<td>Asset change without intent</td>
<td>Insufficient training</td>
<td>6</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-performance of procedures</td>
<td>Non-motivated and dissatisfied staff</td>
<td>8</td>
<td>4</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Incident record

<table>
<thead>
<tr>
<th>Event number</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of incident</td>
<td>Actual / Attempted / Suspected</td>
</tr>
<tr>
<td></td>
<td>Deliberate / Accidental / Error / Not known</td>
</tr>
<tr>
<td>Impacts</td>
<td>Assets affected</td>
</tr>
<tr>
<td></td>
<td>Way of damage, value, cost, losses</td>
</tr>
<tr>
<td>Incident solution</td>
<td>Dates – incident investigation commenced date, incident end date, impact end date, incident investigation completion date</td>
</tr>
<tr>
<td>Person responsible for incident solution</td>
<td>Engineer</td>
</tr>
<tr>
<td>Storage place of records from incident investigation</td>
<td>Incident DB</td>
</tr>
<tr>
<td>Involved person / perpetrator</td>
<td>Description</td>
</tr>
<tr>
<td></td>
<td>Motivation</td>
</tr>
<tr>
<td>Actions taken to resolve incident</td>
<td>Operator monitory</td>
</tr>
<tr>
<td>Actions planned to resolve incident</td>
<td>-</td>
</tr>
<tr>
<td>Other actions (e.g., external)</td>
<td>Test of new cleaning compound</td>
</tr>
<tr>
<td>Conclusion</td>
<td>(major / minor)</td>
</tr>
<tr>
<td>Involved and notified persons / entities</td>
<td>(name, role, date, eventually signature)</td>
</tr>
</tbody>
</table>
5. SUMMARY

Main result of this paper is proposal of critical factors analysis with example of application. Mentioned factors can affect soldering process dependability. The methodology includes analysis of defects risk and also the way of risk management. Information is one of mentioned critical factors of process dependability. By the reason of information importance for soldering process management the risk analysis method was chosen from information technology area.

Other result of this paper is critical factors analysis enlargement by incident management methodology. Incident means a defect of soldering process. This part describes necessary activities for getting incident under control, i.e. causes identification, their elimination and incident impact decreasing.

In this way paper suggests compact methodology approach, which tries not only to prevent defects, but in the case of defect to remove causes as soon as possible, to decrease impact and to obtain important information for next prevention and improvement.

ACKNOWLEDGEMENTS

This paper has been supported by the research plan of Ministry of Education, Youth and Sports of Czech Republic No. MSM4977751310 “Diagnostic of Interactive Processes in Electrical Engineering”.

REFERENCES

A review of finite element analysis of metal cutting

Q. Xu
School of Science and Technology
University of Teesside, Middlesbrough, Tees Valley, TS1 3BA, UK

Z. Lu
School of Computing and Engineering
University of Huddersfield, Huddersfield, HD1 3DH, UK

Abstract

Metal cutting is one of the most widely-used manufacturing processes. Research on this process based on finite element has developed and matured enough to give valuable insight of this process and, sometimes, good agreement with experiment. However, due to the complex nature of the metal cutting process and the non-linear nature of finite element analysis required for this process, the analysis using finite element method of this process is not straightforward. This paper reviews the fundamentals of the finite element simulation of metal cutting. It also includes some significant advancements and applications reported recently. Conclusions and comments on future research are also presented. This paper contributes by helping to establish comprehensive knowledge about the finite element simulation of metal cutting and the state-of-art knowledge in this field.

1 Introduction

The competitive nature of the global economy demands industry to maximize the efficiency and quality, as well as minimise response time and costs in production [1-2].

Metal cutting is one of the most widely-used manufacturing processes to produce the final shape of products, and its latest advancements are paralleled with the developments in materials, computers, and sensors [2].

The final shapes of most mechanical parts are obtained by a series of manufacturing operations. Bulk deformation processes, such as forging and rolling, are mostly followed by a series of metal removing operations in order to achieve parts with desired shapes, dimensions, and quality of surface finish. Though those two types of manufacturing processes are distinct, both rely on the behaviour of material past yield point.

Further more, high speed machining has been developed over the last two/three decades in order to increase productivity. Its advent requires a new understanding of cutting mechanics under such new conditions [1-2].

Due to its importance, a great deal of research has been devoted to understanding the mechanics of this process. The purpose of analysing metal cutting is to understand and predict process variables such as stresses, temperatures and cutting forces for given tool geometrical parameters (rake angle, clearance angle and edge preparation), cutting conditions (feed, cutting speed and depth of cut) and workpiece and tool material properties. To achieve this goal, experimental, mechanistic and analytical techniques are used. In recent years, however, with the surge in computational power and capacity, the focus has turned to numerical simulation of the process through finite element methods. More and more such simulations have appeared in the literature [4-12] and in the proceedings of conferences (such as the CIRP Workshop on FEM Modelling of Machining Operations) [13-14].

This paper has resulted as a preliminary literature review in this topic. It focuses on to the fundamental questions and the key issues to be encountered in the application concerning the finite element simulation of metal cutting process. In addition to that, the authors’ own comments and views will be presented whenever it is pertinent and appropriate. This paper contributes to broad views on the frame-work of computational plasticity and some practical aspects about metal cutting simulation.

2 The main characteristics of metal cutting

The cutting process is a very complex one [1-4]. From a computational mechanics point view, it is characterised by:

1. Large deformation of the material with high strains (1 or higher)
2. High strain rates (10⁴ to 10⁵ s⁻¹ and higher)
3. Thermal-coupled with high temperatures in the range of 200 to 1000°C and higher
4. Complicated friction conditions at the interface
5. Chip incipient and separation
6. Serrated chip or continuous chip depending on process parameters
7. Highly localized heat generation associated with large plastic deformation and high friction
8. Tool wear and vibration in real machining processes

In the machining process the stress, strain, strain rate and temperature variables will all be dependent on the cutting parameters such as the feed rate, and the cutting speed as well as on the geometrical features of the tool such as the rake angle and the nose radius. Thus, from a computational mechanics point of view, machining/cutting process is essentially a thermal-coupled elastic-viscoplastic problem. However, it is often simplified and a rigid-viscoplastic material model is used instead – it gives reasonable results for chip morphology, contact stresses and tool wear etc. while being more CPU efficient. The chip separation complicates this problem further. Thus, the finite element method is a natural choice. In fact, as far as the author’s awareness is concerned the development of a computational method with a finite element technique has been inspired by potential applications in civil, mechanical and manufacturing engineering (such as metalforming and machining) and its rigour has benefited from the input of mathematical knowledge.

3 FE Formulation

The FEM has been used for the simulation of metalworking operations. The majority of these analyses use either the Lagrangian approach or the Eulerian approach.

The Lagrangian approach is a natural choice for solid mechanics. The FE mesh consists of material elements and the computational grid deforms with the material (workpiece and/or tool). This approach is particularly convenient to track unconstrained material flow as the FE mesh presents the material boundaries. The Eulerian approach is more suitable for fluid-flow problems. The mesh consists of spatially fixed grids and the material properties are calculated at fixed spatial locations as the material flows through the mesh. Interested readers can find comprehensive coverage of general computational plasticity in references [15-17].

Critical observations have been made about those approaches with respect to machining process simulation in several publications (say, such as [18]).

1. The Lagrangian approach is relatively easy in formulation and implementation.
2. The analysis can be started from indentation to the incipient stage to steady state. The chip is formed easily to its proper thickness, and no prior assumption is needed about the shape of the chip.
3. The mesh is severely distorted adjacent to the tool tip and re-mesh is needed. Not All the FEA packages have this capability.
4. A separation of nodes from in the front of the tool tip may be needed. Machining can essentially be considered a (shear) extrusion process on a micro-level – material will flow either across or under the cutting tool as in any forging operation.
5. A fracture criteria is only required where the chips break off following formation if the chip is discontinuous (which as in reality will depend on the tool geometry and cutting parameters). This is currently possible with DEFORM using a ‘Damage Criteria’. Reference [3] described this aspect, too.
6. The Eulerian approach is more computationally efficient and is best for modelling the region around the tool tip.
7. It requires a prior assumption of the shape of the chip.
8. An iterative procedure is needed for the convergence of boundaries.
9. History-dependent material properties have to be interpolated in an approximate way and material points have to be traced back to the fixed mesh points.

Consequently, the Arbitrary Lagrangian-Eulerian (ALE) approach has been applied to the machining process recently [18]. This development is inspired mainly by the realisation of the strength of the combination of the Lagrangian and Eulerian methods in modelling some of the characteristics of the machining process. Interested readers can refer to those publications directly (such as [18-19]).

The choice of finite element formulation is a very serious issue if it is intended to develop in-house software. Although it is true that traditionally researchers were mainly using in-house software in their research work (over 70% up to mid 90’s) [8], the use of commercial packages has increased significantly over the last 5 to 7 years. The question for most application researchers who are not intending to develop their in-house software is which software to select and which finite element formulation should be considered.
4 Basic problem

4.1 Material model and flow stress

A realistic material behaviour model is very important and it should be able to depict the main behaviour under large strain, high strain rate, and thermal effects. However, various simplifications have been introduced for specific research application. The well-known Johnson-Cook formulation is often used. The flow stress is given by:

\[
\sigma = (A + B \cdot \dot{\varepsilon}^m) \left[ 1 + C \ln \left( \frac{\dot{\varepsilon}}{\dot{\varepsilon}_0} \right) \right] \left[ 1 - \left( \frac{T - T_{\text{amb}}}{T_{\text{melt}} - T_{\text{amb}}} \right)^\frac{n}{m} \right]
\]

(1)

where \( \dot{\varepsilon} \) is the plastic strain, \( \dot{\varepsilon} \) is the strain rate, \( T \) is the temperature, \( T_{\text{amb}} \) is the room temperature, \( T_{\text{melt}} \) is the melting temperature. \( A, B, C, \) and \( n \) and \( m \) are rheological parameters. Some variation has been proposed in [12]. Recent development in the determination of flow stress highlighted the use of inverse method.

4.2 Friction in chip-tool interface

A typical stress distribution on the rake face is illustrated in Figure 1. It is mathematically presented as:

\[
\tau_f = \mu \sigma_n \quad \text{when } \tau_f \leq k
\]

(2)

\[
\tau_f = k \quad \text{when } \tau_f \geq k
\]

(3)

where \( \tau_f \) is the frictional stress, \( \sigma_n \) is the normal stress, \( \mu \) is the coefficient of friction, and \( k \) is the shear stress of the chip material. Work published in [11] proposed a different model and its effects on residual stress.

4.3 Chip formation

Chip separation from the workpiece and chip formation (either continuous for discontinuous) is one of very important aspects in metal cutting. It was observed that over the past 8 years or so, numerical simulation of machining processes has been the subject of intensive research, in which aspects of discontinuous chip formation and algorithms for element separation have been addressed, but the modelling and analysis of such problems still require further attention [3].

Many criteria have been proposed in the literature, ranging from simple geometrical or strain criteria to more complex fracture-mechanics criteria. Even if the appropriate type of criterion is chosen, there is no clear physical justification as to what critical value should be adopted [18]. This is an area further fundamental research is needed.

Figure 1. A typical stress distribution on the rake
Nevertheless, pertinent to practical application of finite element simulation of the cutting process with commercial software, it has been reported that the Johnson-Cook failure mode has been used with ABQUS/Explicit [8]. The strain at failure was assumed to be function of stress state and temperature as:

\[
\vec{\varepsilon}_f = \left[ d_1 + d_2 \exp\left( \frac{d_3}{q} \right) \right] \left[ 1 + d_4 \ln\left( \frac{\vec{\varepsilon}}{\vec{\varepsilon}_0} \right) \right] \left[ 1 + d_5 \left( \frac{T - T_{\text{trans}}}{T_{\text{melt}} - T_{\text{trans}}} \right) \right]
\]

\[\omega = \sum \frac{\vec{\varepsilon}}{\vec{\varepsilon}_f}\]

where \( \omega \) is the damage parameter and the material is deemed failed when it reaches to unity. \( d_1, d_2, d_3, d_4, \) and \( d_5 \) are the failure constants to be determined by experiment.

It has also reported [20] that a continuum damage based chip breaking model has been developed by the researchers at Scientific Forming technologies Corporation.

5 Available software

The following software has been reported to be able to analyse the metal cutting process. They include:
- ABAQUS/Standard and ABAQUS/Explicit
- ALGOR
- DEFORM 2D AND DEFORM 3D
- FLUENT
- LS DYNA

A detailed comparison of their functionality is desirable but that information is not available yet. In addition to that information about their computational efficiency is scattered in various publications and it is not readily available.

6 Recent Applications

6.1 Applications of Deform 2D/3D

Deform 2D and Deform 3D is special software dedicated to metalworking processes (metalforming and machining). Recently, significant advancements and its applications have been reported (either the Scientific Forming Technologies Corporation [20-22] or other research groups [8, 13-14]). Some of the significant applications are reported in [20-23]:

1. Tool life prediction and comparison. Good agreement between predicted tool life and experimental observation is reported in [21].
2. 3D simulations (drilling and milling). Example is shown by Figure 2.
3. Chip control. Example is shown by Figure 3.

6.2 Other Developments

There are a few new developments worthy mentioning:
1. FEA simulation of chip separation in machining unidirectional glass fiber-reinforced composite [24] where dedicated material damage modes were modeled via ABAQUS user-subroutine;
2. Finite element method and artificial neural network (FEM-ANN) combined to investigate the residual stresses and the optimal cutting conditions [25];
3. Integrating Finite Element and Multi-body Simulation for investigating the machining performance [26];

7 Conclusion
Finite element simulation is the new trend in research to understand metal cutting mechanics. Significant progress has been achieved in both the theoretical framework and application. Valuable insights have been obtained and good agreements with experiment have been reported.

Whilst computer simulation of metal cutting processes has gained reasonable success, certain fundamental issues still need to be addressed by the research community. These issues include the development and then its implementation of more advanced plasticity theory (such as non-local plasticity theory) and physically-based damage mechanics model for chip separation.

![Figure 2. Chip Formation in Drilling [21]](image1)

![Figure 3. Three dimensional simulation of a spiral chip [21]](image2)

**References**


An Intelligent System for Analyzing Welding Defects using Image Retrieval Techniques

Raoul Pascal Pein\textsuperscript{1*}, Joan Lu\textsuperscript{1}, John Birger Stav\textsuperscript{2}, Qiang Xu\textsuperscript{3}, Miro Uran\textsuperscript{4}, Luboš Mráz\textsuperscript{5}

\textsuperscript{1}Department of Informatics, University of Huddersfield, Huddersfield HD1 3DH, UK  
\textsuperscript{2}Fac. of Informatics and e-Learning, Sor-Trøndelag University College, Trondheim, 7004, Norway  
\textsuperscript{3}Technology Futures Institute, University of Teesside, Middlesbrough, TS1 3BA, UK  
\textsuperscript{4}Welding Institute, Ljubljana, 1000, Slovakia  
\textsuperscript{5}Welding Research Institute, Bratislava, 83259, Slovakia

ABSTRACT

The development of new approaches in image processing and retrieval provides several opportunities in supporting different domains. The group of welding engineers frequently needs to conduct visual inspections to assess the quality of welding products. It is investigated, if this process can be supported by different kinds of software. Techniques from a generic CBIR system have been successfully used to cluster welding photographs according to the severeness of visual faults. Similarity algorithms were used to automatically spot faults, such as cracks and gas pores.

INTRODUCTION

The techniques of Content-Based Image Retrieval (CBIR) systems can be applied in various disciplines. Not only image processing in computing areas such as image libraries [1, 2] can be beneficial, but also other application domains. For example, the importance of CBIR is especially increasing in medical applications, where a large amount of X-ray images are produced for diagnosis [3], as well as the engineering sectors, where investigations in 3D model retrieval have been addressed recently [4].

In the field of CBIR, the feature vector paradigm is a commonly used technique [5, 6]. Several different features have been developed to describe the most relevant content of images in a highly condensed way [7, 8]. Instead of comparing the complete images, only this information is used. This allows for fast and considerably accurate CBIR.

In welding industry, Non Destructive Testing (NDT) could produce a huge amount of images in the forms of photos, digital images or X-ray pictures containing different types of welding defects. These images are vital to assess the quality of industrial products. According to the definitions defined by ISO 6520, there are 80 defect types listed. For each type, the defects could be grouped or classified into different degrees of severity. However, advanced image retrieval techniques are rarely applied into the field of welding, particularly for analyzing welding defects taken by X-ray pictures. Traditionally, analyzing X-ray pictures may involve expensive facilities, which is costing and time consuming [9]. Many welding defects in radiographics can be spotted by trained humans, but the sheer amount of produced images makes it difficult to manually analyze everything accurately. This provides a very good opportunity for computer based analysis. The key to success is to find a feature that describes a defect detailed enough to allow automated detection [10]. It follows that research into developing an advanced system for retrieving, analyzing, classifying and recognizing welding defects is of interest in both academic study and industrial applications.

A recent overview concerning the analysis of radiographic weld seam images has been published by da Silva and Domingo, split into image processing [11] and pattern recognition [12] tasks. The mentioned approaches cover several problems from general noise reduction to spotting and measuring the strength of specific welding defects. The research of Felisberto et al. [10] is concerned with the task of detecting defined objects within radiographic images, which is highly valuable information for further analysis. Otherwise, sophisticated and time-consuming

\*Corresponding author: Tel: +44(0)1484 47-2336; E-mail: r.p.pein@hud.ac.uk
fault detection algorithms are prone to extract information from the wrong areas, making it difficult to determine the optimal processing parameters.

The objective of this research is to develop an analysis tool to spot and measure welding defects with minimal effort to the user. The challenge of this research is to find alternative ways to spot defects without being affected by background noise. It is crucial to define classifiers that are able to distinguish between different image areas, such as background, metal sheets, pipes and welding seams. These classifiers can be used to find the regions-of-interest where detailed analysis can be applied. Further, the characteristics of each region are useful to find the best tuning parameters.

METHODS EMPLOYED

In this investigation, methods from the CBIR field are employed to analyze radiographic images of various welding defects. ISO 6520 (Classification of geometric imperfections in metallic materials) is used as supporting document to distinguish different types of welding defects.

- Segmentation: i.e. an image is segmented into regions;
- Extraction: i.e. extract different kinds of feature vectors for each one.

These feature vectors describe the region with highly condensed information. For example, feature vectors may contain the histogram values of a region to describe the color distribution [7]. Wavelet coefficients are used to describe information such as frequencies [8] and shapes [13]. Extracted features can be compared to each other and be used in retrieval. The detailed description of using feature vector algorithms for fault detection is as follows.

The idea is to split the original image into several smaller sections. Each section represents one or more object classes. In the following, the original image is split into 10x10 regions and various feature extraction algorithms are run on each of them. The resulting data set allows for a comparison of each two sections according to particular feature vector characteristics. Dependent on the kind of feature vector used, an average feature may be computable, e.g. calculating the arithmetic or geometric mean of a vector.

An alternative approach based on pure pattern recognition is described in [14]. It performs several steps of image pre-processing on the input image to end up with a gray-scale image, where the relevant areas have considerably different values than the average value. The transformations include techniques such as median, edge detection and Hough transform. That approach is capable of measuring the exact diameter of gas pores and the length and width of cracks. Hence, the success of that solution is dependent on the fine-tuning of various parameters used to remove background noise. Also, overlapping defects cannot always be detected correctly, as it is difficult to detect the boundaries of each single one.

SYSTEM DESIGN

The system design considers several actions and user requirements as follows:

1. Load sample image
2. Segment image
3. Extract feature vectors from segments
4. Select query: either existing segment or specific feature vector
5. Compare each segment to the query and collect similarities
6. Analyze similarity pattern

Most processing is done automatically and the tuning parameters are supposed to be hidden to the user. The automated tasks include: image segmentation, image analysis, comparison of segments, highlighting possible defects. The user interaction is covered by: loading an image, optionally selecting an initial region-of-interest and manually checking the potential defect locations.

The proposed system provides a straightforward user interface. It can load images, which are then segmented and analyzed automatically. The user’s main task is to select a region of interest. Based on the underlying feature
information, the system automatically compares this region to the others. The result is an image, where the similarity to the selected region is displayed. It is assumed, that a suitable feature vector is able to highlight regions with significantly different content, i.e. local defects or other objects.

A completely user-independent way to generate results is the automatic selection of a source region or the use of average feature values. The latter approach is expected to be most powerful in sample images with relatively large homogeneous areas and only a few different spots.

IMPLEMENTATION

The prototype has been developed completely in Java. It uses a couple of external libraries. The basic image processing is done with the help of the Java Imaging Utilities [15]. All CBIR related functionality is directly taken from the image retrieval system developed prior to this research [16].

Internally, the source image is chopped into rectangular segments of a pre-defined size (i.e. 10x10 pixels). For each segment, a feature extraction algorithm is applied and the vector is attached to the corresponding sub image. The generated ad-hoc image repository is stored in the main memory. All further analysis is then handled like a complete CBIR retrieval. The query contains the desired feature vector and all segments are compared to it, resulting in a similarity between 0.0 and 1.0. For each segment, a color is calculated based on the respective similarity. This color is then used in the visualization to show the difference of each segment to the original query.

The feature used consists of 12 stochastic moments of the image histograms [7]. As these values can be interpreted as the scalar parts of a feature vector, it is possible to calculate the mean value of multiple features, resulting in a single average vector. In this case, the simple arithmetic means of each component are calculated and used as a new query.

TESTING

Testing data prepared using two representative cases, i.e. (name), in welding defects. Testing is carried out in the following stages:

1. Collection of 25 digital radiographic sample images
2. Manual acquisition of defects
3. Software supported detection of cracks and gas pores
4. Software supported comparison analysis
5. Assessment of automatic query generation
6. Comparison of results

The reference system is the previously developed detection tool for gas pores and cracks [14]. Provided that the algorithm parameters are set accordingly, the detection works with a reasonable precision. Regions detected in that tool should also be spotted in the proposed tool, which is using histogram-based CBIR algorithms. The samples contain different kinds of defects and are analyzed as described above.

The regions spotted manually and by the two tools are compared according to their calculated significance. Further, the results of the automatic query generation are compared to the previous detection methods.

The testing has been carried out on a Intel Core 2 Duo PC with 2 GHz and 2 GB of main memory. The operating system is Windows XP and the Java environment is the current version 1.6.

RESULTS

Two representative test cases are presented below. For the first one, a sample image with a T-shaped welding seam, showing a crack, is used. It is to measure the ability of spotting a single defect in an inhomogeneous environment. For the second test case, a welding with several cracks and gas pores is chosen. The smooth background is expected to be suitable to calculate a suitable mean value, representing the intact area.
CASE ONE: SINGLE CRACK

In Figure 1, a T shaped welding seam is depicted. There is an obvious crack in the central part running orthogonally from the centre to the second seam. This crack clearly emerges from the corresponding black/white image and can be spotted against the noise. The vertical seam on the right appears as a dotted line.

After processing, the crack itself is enclosed within 4 clusters, while the largest cluster contained in the image (ranking of 168) is located directly on the main part of the crack. The second largest cluster (ranking of 46) is a straight part of the vertical seam. Several small clusters (ranking below 30) are scattered across the whole image.

The corresponding CBIR analysis results are depicted in Figures 2 (a) and 2 (b). A pure red indicates a high, yellow an average and white a very low similarity. The effects at the upper and left border are due to a dark line in the original image and should be ignored.

The left image shows the similarities of each 10x10 region based on a region lying directly on the seam, emphasizing the intact part of the seam in a strong red. At the edges, it drops gradually towards a bright white. In the central area, the colour is less smooth than in the surrounding parts. Several brighter areas can be spotted and the large crack is represented by a bright line. Above the crack, a small bright spot at the upper edge of the seam can be seen.

The right image queries with a segment lying in the centre of the main crack, changing the point-of-view. This does not invert the previous result, but emphasizes the difference of each segment to the fault. In this case, the intact
area of the seam is still closer to the crack than the background material. The image is similar to the filtered edge detection image, emphasizing the crack and the vertical seam.

**CASE TWO: MULTIPLE CRACKS AND GAS PORES**

(a) Crack detection results

(b) Gas pore detection results

(c) Similarity results based on histogram average

Figure 3: Seam with multiple cracks and gas pores
A second example (fig. 3) contains an X-ray image containing both cracks as well as. The welding seam is interrupted by some transversal cracks, which are connected by a longitudinal one. In addition, the material is interspersed with gaseous inclusions of varying diameter.

The reference crack analysis is depicted in fig. 3 (a). Both the horizontal and the vertical cracks are contained in various clusters. Additionally, the edges of the gas pores are highlighted. In the reference gas pore analysis (fig. 3 (b)), all the major circular defects are detected and measured reasonably well. Several of the less prominent pores as well as the cracks are ignored.

The final image (fig. 3(c)) is the result of the proposed CBIR based analysis. As described above, the image is split into several quadratic sections. For each section, a feature vector containing 12 stochastic moments is extracted. The query is then constructed by determining the arithmetic mean of all sections. Sections close to the query are represented in red, a lower similarity by colors gradually changing to yellow and then white. While the majority of the image is red, virtually every defect caused a lower similarity and thus different shade of color. The detected areas-of-interest are the same as detected by the manually fine-tuned detection algorithms.

ANALYSIS AND DISCUSSION

The detailed analysis is carried out on a set of x-ray images of welding seams for two cases: i.e. single crack and multiple cracks and gas pores.

CASE OF SINGLE CRACK

![Image of detected crack](a)

![Image of detected crack](b)

Figure 4: Detail view of detected crack

Table 1: Similarity values in the surrounding area of the crack

<table>
<thead>
<tr>
<th>Similarity values</th>
<th>0.40389</th>
<th>0.46589</th>
<th>0.39409</th>
<th>0.27860</th>
<th>0.58690</th>
<th>0.45240</th>
<th>0.35060</th>
<th>0.50960</th>
<th>0.52321</th>
<th>0.54359</th>
<th>0.27860</th>
<th>0.50960</th>
<th>0.52321</th>
<th>0.55854</th>
<th>0.32538</th>
<th>0.55616</th>
<th>0.50440</th>
<th>0.52425</th>
<th>0.34011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarity values</td>
<td>0.49414</td>
<td>0.59685</td>
<td>0.71192</td>
<td>0.78434</td>
<td>0.59086</td>
<td>0.47259</td>
<td>0.53531</td>
<td>0.29895</td>
<td>0.45841</td>
<td>0.50105</td>
<td>0.34819</td>
<td>0.46589</td>
<td>0.50960</td>
<td>0.52321</td>
<td>0.32538</td>
<td>0.55616</td>
<td>0.50440</td>
<td>0.52425</td>
<td>0.34011</td>
</tr>
<tr>
<td>Similarity values</td>
<td>0.49409</td>
<td>0.50010</td>
<td>0.48028</td>
<td>0.71777</td>
<td>0.30139</td>
<td>0.64748</td>
<td>0.84042</td>
<td>0.98333</td>
<td>0.99049</td>
<td>0.99524</td>
<td>1.00000</td>
<td>0.92962</td>
<td>0.96659</td>
<td>0.81462</td>
<td>0.71557</td>
<td>0.34549</td>
<td>0.46965</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity values</td>
<td>0.59154</td>
<td>0.48241</td>
<td>0.40435</td>
<td>0.39358</td>
<td>0.49494</td>
<td>0.43008</td>
<td>0.49668</td>
<td>0.54293</td>
<td>0.61670</td>
<td>0.16161</td>
<td>0.48525</td>
<td>0.72035</td>
<td>0.61840</td>
<td>0.34550</td>
<td>0.31583</td>
<td>0.81764</td>
<td>0.74736</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity values</td>
<td>0.48450</td>
<td>0.40852</td>
<td>0.44289</td>
<td>0.32219</td>
<td>0.56588</td>
<td>0.36902</td>
<td>0.44273</td>
<td>0.54061</td>
<td>0.57005</td>
<td>0.54026</td>
<td>0.32127</td>
<td>0.45627</td>
<td>0.62968</td>
<td>0.52606</td>
<td>0.50291</td>
<td>0.42079</td>
<td>0.66518</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results for the largest crack are shown in figure 4. The details of the similarity calculations are presented in table 1 and directly correspond to the 10x10 pixel areas and the respectively displayed color. The query segment achieved the highest similarity of 1.0 (bold font). The arbitrary threshold of 0.6 is chosen to underline the segments that are considered to be related to the crack. A comparison of the crack and its surroundings indicates a strong gradient between the two areas. In most cases, the similarity drops by more than 0.3. This gradient is a further indicator to distinguish between real cracks and generally noisy areas such as seen directly above the crack (fig. 2(a)).

Two other small areas in the image have a high similarity value and a strong difference to their neighbors and are located in the upper boundary line of the seam. These areas may either be defects or image errors, which need to be checked by an expert. Interestingly, the same three areas also produced from the result image, where the intact seam is highlighted (fig. 2(a)).

The detected crack is comprehensible when comparing the results with the original image. Regarding that the detail level of the original image is not perfect and some noise clutters the background, the achieved accuracy seems to be reasonable. The majority of intact areas achieved similarity values below 0.6 and can be ignored in further processing steps. In the other cases, a closer investigation with a different analysis approach should be applied to
classify the segments in detail. In the example presented, the vertical seam is difficult to distinguish from the actual crack, because the results are relatively high.

The results of the other detection approach are quite similar. The main crack has been detected and has been ranked with the highest significance. Other features such as the vertical seam or other suspicious areas have also been found. Yet, further experiments indicated, that the similarity based approach is more robust in varying image conditions. Finding the best parameters for the edge detection algorithm is not trivial and needs to be adjusted for each input image. The approach proposed in this paper only requires an initial query section in order to produce meaningful results, as it spots relative differences, not absolute ones.

Figure 5: Similarity values in the surrounding area of the crack.

The comparison between the manually assessed sections and the measured similarity results in the surrounding area of the crack is shown in figure 5. Each line stands for a single row. The top row is numbered 1 and the bottom one 5. The x axis describes the width of the area from left (1) to right (17). The similarity values are displayed on the y axis. Two thresholds are set to define the membership between fault (>0.65) and intact (<0.55) area. The space in between represents the area of uncertainty, where no definite conclusion is possible.

The original image is manually assessed and each section is ranked as “faulty”, “intact” or “uncertain”. The system ranked 69 out of 85 sections identical to the human assessment. In 14 cases, it was unclear either for the human or the system, if a segment is faulty or not (highlighted by circles). In two cases, the automatic result deviates from the human opinion (highlighted by squares).

### CASE OF MULTIPLE CRACKS AND GASPORES

Table 2: Average feature vector

<table>
<thead>
<tr>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Green</td>
<td>Blue</td>
<td>Red/Green</td>
</tr>
<tr>
<td>77.2371</td>
<td>82.2430</td>
<td>33.2856</td>
<td>28.8170</td>
</tr>
</tbody>
</table>

Using the average as query (table 2), a high similarity can be interpreted as a good material quality. Defects can be spotted by searching for low similarities. The background achieved results above 0.8 while the obvious defects (i.e. all defects also spotted by the reference algorithms) are ranked below 0.5.

Both approaches have some difficulty to detect the less prominent parts of the longitudinal crack. In both cases, the threshold between good quality sections and defect sections cannot be set to a distinct value, without losing potentially relevant information. Again, the neighboring regions could be the key for the final decision.

### SUMMARY

The single crack case proved its ability to highlight areas of interest without the need of highly specialized algorithms. The results are helpful in drawing the user’s attention to faulty areas. Comparing the manual assessment to the threshold based one, only two segments out of 85 are ranked false. In the second case, the use of the average feature vector value for the ranking seems to be a promising approach. Mainly intact samples should lead to a very even distribution of similarities. Yet, the algorithm cannot recognize and ignore unimportant background areas yet.
CONCLUSION AND FUTURE WORK

The proposed methods are easily to be used to cluster radiographic images into different regions-of-interest. It is shown that relevant regions, such as welding seams or cracks can be selected manually to highlight related areas.

Also, automatic detection of welding defects is achieved, i.e., an automatic determination of the average query is capable of highlighting regions that deviate from the background, i.e., defects. This may significantly help welding industry in NDT to improve trainees abilities to spot welding defects in training scenarios.

Further research is required to determine the efficiency of other feature vectors. The current work focused on 12 stochastic moments of histograms. Other features may be more efficient in spotting defects, e.g., wavelet-based ones [8].

ACKNOWLEDGEMENTS

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The authors deeply appreciate the X-ray samples provided by Miro Uran, Welding Institute, Ljubljana, Slovenia and Luboš Mráz, Welding Research Institute, Bratislava, Slovakia.

REFERENCES

[14] Raoul Pascal Pein, Joan Lu and John Birger Stav, An Investigation in Image Retrieval for Analysing Welding Defects, Accepted at Infosci09, Leipzig, April 2009
A methodology for the determination of foamed polymer contraction rates for the cryogenic manufacture of shoe insoles

V.G. Dhokia, P. Crabtree, S.T. Newman, M.P. Ansell
IdMRC Department of Mechanical Engineering
University of Bath,
Bath, BA2 7AY, UK

ABSTRACT

The ability to produce products, suited to a particular individual is becoming more prevalent in today’s society and because of this more efficient and rapid methods for manufacturing are required. One such method being currently developed is the cryogenic machining of soft materials for producing personalised insoles and outsoles. One of the major elements of cryogenic machining is the freezing of soft polymers, which will subsequently contract. This paper describes a method for predicting and compensating for this cryogenic material contraction effect for a commonly used shoe midsole foamed polymer, namely Ethylene Vinyl Acetate (EVA). Using the linear coefficient of thermal expansion a scaling factor for EVA has been developed for its cryogenic machining. This factor is then applied to the X, Y and Z scaling within the CAM software. The process is tested for a series of EVA cube test pieces and the results provide a scaling factor which shows that the linear scaled dimension are within 1% of the measured contracted X, Y and Z dimensions.

1. INTRODUCTION

Ethylene Vinyl Acetate (EVA) is an extensively used material for a number of applications, particularly for shoe soles, mid soles and insoles [1-3]. The material is durable and offers substantial cushioning and comfort to the user. EVA products including shoe soles are formed using traditional injection moulding technology [4], which is extensively used in the mass produced market.

However the ability to produce personalised shoe soles and insoles using the current state of the art injection moulding technology is not possible due to the complexities and high cost of producing individual moulds for individual designs. A concept currently being pioneered at the University of Bath is the cryogenic machining of soft materials, in particular EVA [5, 6]. However one of the major problems of the cryogenic machining of such materials is the contraction of the material after it has been subject to a cryogen, such as liquid nitrogen (LN$_2$). This contraction can distort the final product shape, thus needs to be compensated before the cryogenic machining process.

This paper reports on the development of a method for establishing contraction factors for EVA at cryogenic temperatures. Experiments are also developed and conducted to provide real time analysis. Finally a framework for using the developed knowledge is proposed and some experimental analysis is provided.

2. EVA CHARACTERISTICS

To understand the behaviour of a material it is important to understand its structure. EVA polymers are produced by copolymerising ethylene with vinyl acetate molecules; the ethylene provides a chain backbone which incorporates a percentage of the monomer vinyl acetate (VA). The properties of EVA co-polymers are strongly dependent on the ratio of vinyl acetate to ethylene in weight [7], the amount of VA added to the ethylene can vary from 2-50% [4]. This percentage content controls the material’s crystallinity and thus its flexibility. There are a number of different grades of EVA that are used in the sports shoe sole industry, particularly for the mid sole as it provides cushioning impact between the running surface and the foot [2]. During the cryogenic process the foamed polymers are subjected to temperatures below their glass transition temperatures. Below this temperature, the amorphous component of the polymer is glassy and when loaded, will distort elastically before yielding or fracturing. A polymer foam in this state is classed ‘rigid’[8].
Figure 1a and b shows a Scanning Electron Microscope (SEM) image of an open celled EVA foam, typical of that used in the midsoles of running shoes. The structure is suitable for sport shoe soles and midsoles as the material will deform on loading and return to its original shape after loading, absorbing energy and limiting the peak force in the heel strike [2].

![Figure 1a: Sports shoe example](image1a.jpg)  ![Figure 1b: SEM image shoe sole EVA](image1b.jpg)

3. CRYOGENIC MACHINING

The traditional method for manufacturing polymer products including EVA is through injection moulding methods [4]. However as stated previously this method is not economical nor is it rapid when considering the production of personalised products, for which there is an ever growing consumer market. The following figure illustrates the traditional mass consumer model for mass producing the same product. This model is extensively used as it has significant cost break benefits over a long period of manufacture for a repetitive mass produced product.

![Figure 2: Traditional mass produced sole](image2.jpg)

Figure 3 provides the author’s vision for the future of producing personalised shoe soles. The major concept is to cryogenically freeze the soft material, in this case EVA, and directly machine it using conventional CNC machines to produce personalised shoe soles and insoles. The process chain is significantly reduced when compared with the traditional mass produced method as shown in figure 2. However at present the cryogenic method is not intended for the mass consumer market, but for the personalised medical market and professional sports domain, for which there are numerous potential applications.
4. DESIGN OF EXPERIMENTS

One of the major challenges for this research work is determining the contraction of EVA due to the cryogenic process as this will dictate the final size of the machined part. There have been 0 [9] conducted on the expansion of EVA during the injection moulding process. It was shown that over a certain size of product the expansion could be controlled, however it was more difficult to determine a uniform method, which could be applied to any type of geometry. This work involves analysing the opposite effect of this, namely contraction. There is however limited knowledge on the behaviour of EVA when subjected to liquid nitrogen.

4.1 TEST PART ANALYSIS

Two grades of EVA were used for testing and analysis. The materials were cut into identical cubes of 25mm³ and measured in the X, Y and Z direction at room temperature using a venire scale. The samples were then placed inside a bath of liquid nitrogen for 1 minute before the dimensions were measured again. Figure 4 illustrates the method used for determining appropriate material dependent contraction factors.

The following data represents the results of the experimental work. Figure 5 shows the height, width and length at room temperature state and contraction state of the EVA material. The major problem with measuring during the frozen state is the rapid increase in temperature of the part when it is removed from the bath. Due to this increase in temperature it was not possible to measure material sections, subsequently only the central sections of the test parts were measured for the height, width and length. However from these early test results it is clear to see a distinct pattern in terms of contraction as shown in figure 5.

Figure 5a illustrates the contraction in the X direction for 5 samples of EVA grade 1 material. The blue line represents the room temperature measured part and the pink line represents the cryogenically frozen measured part. Figure 5b and 5c shows the difference in the Y length and Z widths respectively. From these plotted results an approximation for the percentage change can be observed for the length X, width Y and height Z.
4.2 DETERMINATION OF CONTRACTION

The test piece results indicate a defined pattern for EVA contraction when subjected to cryogenic temperatures through the use of liquid nitrogen. From the obtained results it was shown that the contraction of the 25mm³ cubes of EVA grade 1 was determined as follows using the principle of linear thermal expansion.

The coefficient of thermal expansion for EVA grade 1 can be calculated using the equation 1.

\[
\alpha = \frac{\partial l}{l \times \partial t} \quad \ldots \ldots 1
\]

Where \( \alpha \) = Thermal expansion coefficient for a given material.

\( l \) = Length

\( t \) = Temperature

The linear dimensional change in x, y and z can be calculated using equation 2. The linear dimensional change is the amount the material contracts in a particular direction, in this case the length X. This can be applied to the Y width and Z height.

\[
\partial x = x \alpha \partial t \quad \ldots \ldots 2
\]

The X length after contraction can be calculated by subtracting the original X length from the change in X length \( (\partial x) \). The predicted contraction for X, Y and Z are shown in figure 6.

Figure 6a illustrates the measured contracted X value and the predicted contracted value. Figures 6b and 6c show this for Y and Z respectively.
The scaling factor for the x length = predicted value for x after contraction / original x length before contraction:

\[
\left(\frac{x - \partial x'}{x}\right) \quad \text{……..3}
\]

Figure 7 expresses the contraction scaling process. The coefficient of thermal expansion for a given material is determined and the linear scaling factors are used to adjust the three linear dimensions (X, Y and Z). The developed cryogenic scaling factor is used to adjust the shoe sole model. The scaled shoe sole is then cryogenically machined using the cryogenic facility.

This process determines scaling factors for linear X, Y and Z. These values can be applied to the model in the CAD environment to adjust the model according to the material to be cryogenically machined. Figure 8 demonstrates the different stages involved in the contraction-scaling factor testing methodology. The first stage
consists of straight line machining of slots. The calculated scaling factor is applied to the straight-line slots and then adjusted accordingly. The part is then machined using the developed cryogenic facility. The machined part is then compared with the original CAD model dimensions. Stage 2 investigates contoured parts and stage 3 investigates complex parts, which is a combination of stage 1 and stage 2. Stage 4 applies the gained knowledge from the previous stages and applies it to the cryogenic machining of a shoe sole.

4.3 APPLICATION OF THE CRYOGENIC FACTOR

A 25mm³ test cube was cut and was measured using a veneer scale and found to have the following dimension, X 25.8, Y 25.33, Z 25.07mm. After the cryogenic freezing process the cube was measured again and was found to have the following dimensions, X 25.37, Y 25.3, Z 24.75.

The coefficient of thermal expansion for this material is calculated:

\[
\alpha = \frac{\Delta l}{l \times \Delta t} = \frac{25.8 - 25.37}{25.8 \times (293.15 - 77.15)} = 7.71605^{-5}
\]

Four other samples were measured and the thermal coefficient for each sample is calculated and the average is calculated to being: 6.30419⁻⁵

The change in X is calculated based on the thermal coefficient of thermal expansion.

\[
\Delta x = x \alpha \Delta t = 25.8 \times 6.30419^{-5} \times (293.15 - 77.15) = 0.351319785mm
\]

Using this value the predicted change in x can be calculated:
\[ x - \Delta x = 25.8 - 0.351319785 = 25.44868022 \text{mm} \]

The cryogenic scaling factor for this particular material is then calculated:

\[ \frac{x - \Delta x}{x} = \frac{25.44868022}{25.8} = 0.986382954 \approx 0.99 \]

This value can be applied to the X, Y and Z scaling function within CAD/CAM software to determine the scaled model geometry.

5. DISCUSSION

Developing a method for determining the contraction of EVA and potentially other soft materials will prove significantly beneficial for the cryogenic machining process. Unlike with other manufacturing methods, the manufacture of shoe soles relies on relatively open tolerances within the region of ± 2mm. A series of experiments were conducted on low density EVA, which is an ideal candidate material for shoe soles and orthotic insoles.

From this preliminary study there appears to be an approximate uniform contraction of the EVA across the materials cross sectional area based on the X, Y and Z directions. The calculated cryogenic scaling factor was applied to a test piece to compensate for the materials contraction during machining. The predicted contraction value showed a 1% difference from the measured contracted value, which is deemed to be a suitable error margin. Further tests using different geometrical attributes and features will indicate the efficacy of such an approach for a wider range of low-density polymer materials.

6. CONCLUSIONS

This study illustrates one of the first attempts at determining contraction factors for EVA subjected to cryogenic temperatures. The initial results have been applied to simple slot milling test pieces and have been found to compensate for the contraction of the EVA when cryogenically machined. Following the initial research into EVA contraction the following conclusions can be drawn:

- A methodology has been defined to develop contraction factors for EVA.
- Test pieces have been identified to test the methodology.
- After the initial testing contraction appears to occur uniformly in X, Y and Z.
- The initial contraction factors for EVA were applied to a cryogenically machined test piece. The test part was compared with the original models dimensions and was shown to be within a specified part tolerance, with a 1% contraction error.

Further work will investigate the contraction of other soft shoe sole materials to demonstrate the viability of such a process. This research provides one of the key drivers for the wider scope of cryogenic machining of soft materials currently being pioneered at the University of Bath.

ACKNOWLEDGMENTS

The work reported in this paper has been supported by a number of grants for Engineering and Physical Sciences Research Council (EPSRC), involving a large number of industrial collaborators. In particular, current research is being undertaken as part of the EPSRC Innovative design and Manufacturing Research Centre at the University of Bath (reference GR/R67507/01). The authors gratefully express their thanks for the advice and support of all concerned.

REFERENCES

1. Even-Tzur, N., et al., Role of EVA viscoelastic properties in the protective performance of a sport shoe: Computational


A power consumption model for slot generation with a CNC milling machine

R. Imani Asrai, S. T. Newman & A. Nassehi
Department of Mechanical Engineering
University of Bath, BA2 7AY, Bath, UK

Abstract
Machining processes, are one of the major activities in manufacturing industries and are responsible for a significant portion of the total consumed energy in this sector. Therefore, performing these processes with better energy efficiency, will significantly reduce the total industrial consumption of energy. In this paper, a mathematical model has been presented for the power consumed by a milling machine while performing a process of cutting a straight slot of constant depth out of a metal workpiece. This model describes the connection between the total power consumed in machining to the process parameters, the workpiece material properties, the cutting tool’s characteristics and the machine’s properties. The result is a function which gives the power consumed by the machine in terms of the above parameters. This model- after being generalised to include the various processes on a machine and taking the non-cutting power-consuming activities into account- will make it possible to identify the amount of energy that is consumed by a machine performing a certain planned machining process without actually performing the process. This, consequently, allows a process planner to integrate the total amount of energy consumed by a specific process plan into the planning process as a criterion. Moreover, the constants of the equation, which represent the machine’s behaviour from the energy consumption point of view, allow comparisons between different machines.

1. Introduction

Growing social awareness and scientific knowledge of the vast impact of human beings’ high consumption of energy on the environment have been increasingly forcing the regulatory bodies to encourage reduction of energy consumption in different sectors by different methods, ranging from taxing over the energy consumption itself to introducing CO_2 emission allowance for large industrial consumers [1]. These regulations along with the high price of energy itself have provided a powerful incentive for research around the methods of reduction in energy consumption, especially in the highest consuming sectors.

In 2006, the total energy used by end users in the UK has been equivalent to 157947 thousand tonnes of oil, of which, 32766 thousand tonnes have been used by industry [2]. “Machinery and equipment”, consuming equivalent to 802 thousand tonnes of oil, has been responsible for 2.45 percent of industrial consumption and more than 0.50 percent of the total energy consumption of the UK. This, in terms of the early 2009’s low US$40 per barrel price of oil, has cost about £140m to the nation. The total amount paid by the factories, according to the added price due to taxing, has been much higher. Machining processes, therefore, being a major activity among the different processes in “Machinery and equipment” category, present a great opportunity for research in the area of energy consumption reduction.

The focus of this paper is on identification of the power used by a machine- a milling machine in particular- as a function of the process parameters by which the machine performs the machining job, by modelling the flow of power delivered to a machine tool into its different elements and recognising the behaviour of power consumed by each component as a function of process parameters. This subsequently would enable a process planner to integrate the total amount of energy used by a specific machine to perform a particular machining job, into the planning method as a new criterion along with the other criteria previously used. The process for which the total power consumption of the milling machine is investigated is the cutting of a straight slot of constant depth out of a metal. Using such limited process conditions leads to simplicity of the model and its verification process, but restricts the model to a very limited extent of applications. However, the methodology used in this paper can be extended to construct a more general model including different cutting processes and also to the power consumption of a machine tool at the times when it is not actually cutting, without major difficulty.

Models for dynamics of a milling machine have been previously studied extensively [3]. These models provide a framework for monitoring and controlling the conditions of cutting processes. One of the key factors of a machine tool’s dynamics, which is also essential in the model presented in this paper, is the force between
the tool and the material being cut. This force, depending on many different parameters ranging from process parameters to tool and material characteristics and the geometry of tool-material contact, has been extensively studied for many different cutting conditions [4]. However, a simple model which suggests a simple form for this force by suggesting a dimensionless constant is applied in this paper. This constant depends on the geometry of tool-material contact and the mechanism of fracture in the material being cut [3, 5].

Recently, studies on the profile of energy consumption in machine tools and the sources of loss in the machining processes are taking place [6]. These studies show that, normally, only 25% of the power delivered to a machine tool is actually used for the cutting process and 75% of that is lost in different ways by the machine tool [6].

The study of energy consumption in systems is usually done by breaking the system into smaller systems and subsystems which makes it possible to compare different energy consuming subsystems. This methodology has been applied to the study of the UK’s energy consumption by studying different sectors such as the transportation, industry, domestic and service. Moreover, each of these sectors are being studied by breaking them into subsectors [2]. The same method is applied to a machine tool as an energy consuming system in this paper.

2. A POWER CONSUMPTION MODEL FOR A MILLING MACHINE CUTTING A SLOT

The case studied in this paper is a milling machine cutting a straight slot of constant depth of \( h \), out of a metal workpiece of shear strength \( r \), by an \( N \) cutting teeth tool of radius \( r \), with spindle angular velocity of \( \omega \) and feed- per cutting tooth- of \( f \). The question is how the total power used by the machine during such a process depends on the parameters above.

The power delivered to the machine, during the time it is cutting a straight slot of constant depth out of metal workpiece, is called \( P \) here. This schema for the power consumption model is illustrated in figure 1, and this power can be used in the following ways;

- Dissipation in electrical resistance of internal circuits of the machine, \( P_R \)
- Dissipation inside the machine in bearings, gears and so on due to internal frictions, \( P_f \)
- Work done by tool on the workpiece, \( \dot{W} \)
- Increasing the kinetic energy of moving elements, \( \dot{K} \)
- Power consumption of other mechanical devices like fans, coolant pump, etc., \( P_e \)

Therefore:

\[
P = P_R + P_f + \dot{W} + \dot{K} + P_e \tag{2-1}\]

The next step is to try to identify the correct representation of any of the above terms.
2.1. ELECTRIC RESISTIVE DISSIPATION, $P_R$

Some of the electric dissipation takes place in linear circuits, such as rotor coils, and is named here as $P_{RI}$. If the equivalent resistance of the machine, as a black box with end points at the power supply points, is $R$, the dissipative power of the internal circuits will be:

$$P_{RI} = RI^2$$  \hspace{1cm} (2.1-1)

where, $I$ is that fraction of the current entering the machine from the web, which goes into the linear circuits of the machine, at any time.

The rest of the electric dissipation happens in non-linear circuits, like the machine’s computer, which will be called $P_{Rel}$ hereafter. If the voltage of the $i^{th}$ device is $V_{nl,i}$ and the current of the $i^{th}$ device is $I_{nl,i}$, the power would be:

$$P_{Rel} = \sum_j V_{nl,j} I_{nl,j}$$  \hspace{1cm} (2.1-2)

Since the voltages of the non-linear devices are usually constant, this is the current which determines the power. Moreover, the current is almost unchanged for most of the time that the machine is on. Thus it is reasonable to assume that the whole non-linear electric power consumption of a machine is a constant.

The total electric resistive dissipation, as a result, can be written as:

$$P_R = RI^2 + \sum_j V_{nl,j} I_{nl,j}$$  \hspace{1cm} (2.1-3)

2.2. FRICTION DISSIPATION, $P_f$

2.2.1. FRICTION DISSIPATION OF THE MACHINE ELEMENTS WHOSE SPEEDS ARE PROPORTIONAL TO THE SPINDLE’S SPEED

There are some machine elements whose motion is linked linearly to the spindle’s motion. The electromotor which turns the spindle and any gears and bearings along the way from that electromotor to the spindle are such elements. All these moving elements dissipate some power through friction whose magnitude is a function of their speed. Since their speed is proportional to the spindle’s speed, the power dissipated by these elements via friction is a function of the spindle’s speed.

Frictional forces, and therefore torques, are from three different origins. Solid or dry friction torque $M_{fd}$, is caused by solid surfaces in contact with each other with relative tangential movement. This is almost independent of the relative speed of motion:

$$M_{fd} = cte$$  \hspace{1cm} (2.2.1-1)

Here, $cte$ means it is independent of speed, but is not necessarily independent of all parameters. Thus;

$$\frac{\partial M_{fd}}{\partial \omega} = 0$$  \hspace{1cm} (2.2.1-2)

Low Reynolds wet friction, $M_{fwl}$, is produced by the laminar motion of lubricants in between the moving machine elements, mostly in fluid bearings, and is proportional to the relative speed of the two solid walls which is itself proportional to the rotational speed of the machine, $\omega$:

$$M_{fwl} \propto \omega$$  \hspace{1cm} (2.2.1-3)

There are a number of machine moving elements such as the machine’s table, whose speed, and therefore the relative speed in its bearings, are not related to the spindle speed, $\omega$. In these cases, $\omega$ must be replaced by the element’s (table) speed. For instance, the table’s speed is actually proportional to the product of speed by the feed rate.

High Reynolds wet friction, $M_{fwh}$, which is from the same origin as the previous one but in the turbulent regime and also the air resistance against the moving elements, is proportional to the second power of the relative speed of the moving elements and therefore the spindle speed:
So the total internal friction torque, \( M_f \), can be written as:

\[ M_f = M_{fd} + M_{fsd} + M_{fsh} \]

\[ \Rightarrow M_f = A_1 + A_2 \omega + A_3 \omega^2 \]  

The power consumed by these frictional torques, \( P_f \), will therefore be:

\[ P_f = \omega M_f \]

\[ \Rightarrow P_f = A_1 \omega + A_2 \omega^2 + A_3 \omega^3 \]

The \( A_i \) coefficients above can actually be a function of the net force and torque exerted on the tool by the workpiece while in operation. The force can change the loads on bearings inside the machine, and therefore, their frictional torque. The torque on the tool changes the torques of the gears inside the machine and their frictional torque. Since the geometry of the machine is almost unchanged during the process, except for the moving table, the loads on the bearings and gears due to the force and torque on the tool are linear functions of these torque and force, more precisely speaking, it’s components. So:

\[ A_i = B_i + \sum_{j=1}^{3} C_{ij} F_j + D_i M_{mod} \]

Where \( F_j \) and \( M_{mod} \) are the \( j^{th} \) component of the net force vector \( \vec{F} \) and the torque exerted on the tool by the workpiece respectively. This equation can be reduced to a simpler form for every type of machine. This is done for a milling machine in section 2.3.3.

### 2.2.2. Friction Dissipation of the Machine Elements Whose Speeds Are Proportional to the Machine Table

The same argument, leads to a similar equation for the dissipation in parts moving with the machine table. However, this is the table speed, \( V_{tab} \), which plays the same role as the angular velocity of the spindle now.

\[ P_{tab} = \sum_{i=1}^{3} G_i V_{tab}^i \]

### 2.3. The Work Performed on the Workpiece by the Machine, \( \dot{W} \)

Finding the mean force between the tool and the workpiece is essential for calculating the torque needed to turn the tool and, therefore, the work done on the workpiece by the tool.

#### 2.3.1. The Force Between the Tool and the Workpiece

The forces shown in figure 2 are those exerted by the cutting tool on the workpiece with \( \vec{F}_T \) and \( \vec{F}_R \) being the tangential and radial components. For such a cutting tool, with depth of cut \( d \) and width of cut \( b \) cutting through a metal with shear strength \( \tau \), the magnitude of the tangential component is given by [3]:

\[ F_T = kbd \tau \]

\( k \), the specific cutting stiffness divided by \( \tau \), is a constant which depends on the geometry of the tool and the workpiece contact, material properties, the friction coefficient on the tool and the material contact surface and the cutting mechanism. This is usually of the order of magnitude of 1. The magnitude of the radial component, \( F_R \), is, by a good approximation, proportional to the tangential component, \( F_T \) [3]:

\[ F_R = \mu F_T \]

with \( \mu \) being a constant which can be either positive or negative.
To calculate the force and torque between the tool and the workpiece, consider the radius of the tool \( r \), the feed per cutting edge \( a \), the depth of cut \( h \) and the angular velocity of the tool \( \omega \) and the number of cutting edges \( N \). Every point on the tool including the tip moves on a cycloid. As the table speed is much smaller than its turning linear velocity this cycloid can be approximated by a circle. From the geometry shown in figure 2, it can be deduced that:

\[
d = a \sin(\theta)
\]  

Using this and the above equations for \( F \), the following can be concluded:

\[
F_T(\theta) = kha \tau \sin(\theta), \text{ for } 0(\theta/\pi) 
\]  

\[
F_R(\theta) = \mu kha \tau \sin(\theta) 
\]  

\[
F_x(\theta) = F_T(\theta) \sin(\theta) + F_R(\theta) \cos(\theta) 
\]  

\[
\Rightarrow F_x = kha \tau \left[ \sin^2(\theta) + \mu \sin(\theta) \cos(\theta) \right]
\]  

and

\[
F_y = F_T \cos(\theta) - F_R \sin(\theta) 
\]  

\[
\Rightarrow F_y = kha \tau \left[ \sin(\theta) \cos(\theta) - \mu \cos^2(\theta) \right]
\]

Averaging these two forces over one revolution of the tool:

\[
\bar{F}_x = \frac{1}{2\pi} \int_0^{\pi} F_x(\theta) d\theta
\]  

\[
\Rightarrow \bar{F}_x = -\frac{kha \tau}{4}
\]  

\[
\bar{F}_y = \frac{1}{2\pi} \int_0^{\pi} F_y(\theta) d\theta
\]  

\[
\Rightarrow \bar{F}_y = -\frac{\mu}{4} kha \tau
\]
Since $a << r$ in most cases, for the torque exerted on the tool, it can be inferred that:

$$M = rF_T = khar \tau \sin(\theta)$$  \hspace{1cm} (2.3.1-14)

Averaging this torque over one revolution of the tool:

$$\bar{M} = \frac{1}{2\pi} \int_0^{\pi} M(\theta) d\theta$$  \hspace{1cm} (2.3.1-15)

$$\Rightarrow \bar{M} = \frac{1}{\pi} kha \tau$$  \hspace{1cm} (2.3.1-16)

Finally, having $N$ cutting edges on the tool:

$$\bar{F}_{tool} = \frac{N}{4} kha \tau \left( \dot{i} - \dot{\mu} \right)$$  \hspace{1cm} (2.3.1-17)

$$M_{tool} = \frac{N}{\pi} kha \tau$$  \hspace{1cm} (2.3.1-18)

The force exerted on the work can, thus, be written in terms of the torque exerted on the work by tool as:

$$\bar{F}_{tool} = \frac{\pi M_{tool}}{4r} \left( \dot{i} - \dot{\mu} \right)$$  \hspace{1cm} (2.3.1-19)

### 2.3.2. Calculation of the Work Done on the Workpiece by the Machine, $\dot{W}$

Having the torque exerted on the workpiece by the tool, the rate of work done on the workpiece by tool can be written as:

$$\dot{W}^* = M_{tool} \omega$$  \hspace{1cm} (2.3.2-1)

$$\Rightarrow \dot{W}^* = \frac{N}{\pi} kha \tau \omega$$  \hspace{1cm} (2.3.2-2)

There is also other work which is performed by the motor pushing the table in the Y direction towards the tool.

$$W^k = \bar{F}_{tab} \bar{V}_{tab}$$  \hspace{1cm} (2.3.2-3)

The force exerted by work on the tool is opposite to the force exerted on the work by the tool. Since the table velocity is in the Y direction, the above equation can be written as:

$$\dot{W}^k = -F_{tool} \bar{V}_{tab}$$  \hspace{1cm} (2.3.2-4)

$$\Rightarrow \dot{W}^k = -\frac{\pi \mu M_{tool}}{4r} \bar{V}_{tab}$$  \hspace{1cm} (2.3.2-5)

The table velocity itself is:

$$\bar{V}_{tab} = -Na_0 \dot{j}$$  \hspace{1cm} (2.3.2-6)

Therefore:

$$\Rightarrow \dot{W}^k = -\frac{\pi \mu M_{tool}}{4r} Na_0 \dot{\omega}$$  \hspace{1cm} (2.3.2-7)
So, the total mean rate of work being done on the tool-work contact can be found by adding the equations (2.3.2-1) and (2.3.2-7):

$$W = M_{tool} \omega \left( 1 - \frac{\pi \mu N \alpha}{4r} \right)$$  \hspace{1cm} (2.3.2-8)

### 2.3.3. Friction Dissipation Redux

Now, equipped with the force and torque of the tool, equations (2.3.1-17) and (2.3.1-18), (2.2.1-9), which gives the power dissipated by internal frictions for the spindle and the elements moving with it, can be rewritten. Since the tool force has only two non-zero components, $X$ and $Y$, the equation (2.2.1-9) can be rewritten as:

$$A_i = B_i + C_{ix} F_{tool,x} + C_{iy} F_{tool,y} + D_i M_{tool}$$  \hspace{1cm} (2.3.3-1)

Equation (2.3.1-19) suggests a linear relationship between the tool force vector’s components and the tool torque. So by defining:

$$E_i = \frac{\pi}{4r} \left( C_{ix} - \mu C_{iy} \right) + D_i$$  \hspace{1cm} (2.3.3-4)

This can be written:

$$A_i = B_i + E_i M_{tool}$$  \hspace{1cm} (2.3.3-5)

The friction dissipation function for the spindle and the elements moving with the spindle can be rewritten:

$$P_{fr} = \sum_{i=1}^{3} \left( B_i + E_i M_{tool} \right) \omega^i$$  \hspace{1cm} (2.3.3-6)

And finally, replacing $M_{tool}$ by its equivalent gives:

$$P_{fr} = \sum_{i=1}^{3} \left[ B_i + E_i \left( \frac{N}{\pi} k_ha \tau \right) \right] \omega^i$$  \hspace{1cm} (2.3.3-7)

The dissipation in the elements moving with the table is given in (2.2.2-1). By replacing the table speed with its equivalent:

$$V_{tab} = N \alpha \omega$$  \hspace{1cm} (2.3.3-8)

It can be concluded:

$$P_{f_{tab}} = \sum_{i=1}^{3} G_i \left( N \alpha \omega \right)^i$$  \hspace{1cm} (2.3.3-9)

Thus, the total friction dissipation function can be found by the sum of (2.3.3-7) and (2.3.3-9):

$$P_f = \sum_{i=1}^{3} \left[ B_i + E_i \left( \frac{N}{\pi} k_ha \tau \right) \right] \omega^i + \sum_{i=1}^{3} G_i \left( N \alpha \omega \right)^i$$  \hspace{1cm} (2.3.3-10)

$$P_f = \sum_{i=1}^{3} \left[ B_i + E_i \left( \frac{N}{\pi} k_ha \tau \right) + G_i \left( N \alpha \omega \right)^i \right] \omega^i$$  \hspace{1cm} (2.3.3-11)

### 2.3.4. Rate of Increase in the Kinetic Energy of the Moving Elements, $\dot{K}$

The power consumed to increase the kinetic energy in the machine’s moving elements, gets dissipated either via a number of sources of friction or turns into work performed on the workpiece. Therefore, although being responsible for a considerable fraction of the machine’s total instant power consumption at times, there is no need to take this into account while studying the mean power consumption of the machine. Moreover, during the process which is studied in this paper, the linear and angular velocities of moving and turning elements are almost constant and there is no significant change in their kinetic energies.
2.3.5. Power Consumption of Other Mechanical Devices Like Fans, Coolant Pump, etc., $P_e$

There are some devices, like fans and the coolant pump, which work persistently during the process having their consumed power unaffected by the process. Therefore, the total power consumed by these devices can be assumed as a constant $P_e$.

2.3.6. The Total Power Consumption Function for a Milling Machine

From the equation (2-1), and using the results of the previous sections for the occurring terms, it can be concluded that the total power consumption of a milling machine must depend on the process parameters and machine parameters in such a form:

$$ P = R I^2 + \sum_j V_{nl,j} I_{nl,j} + \sum_{i=1}^3 \left[ B_i + E_i \left( \frac{N}{\pi} k h a \pi \right) + G_i \{ Na \} \right] \omega^i + \frac{N}{\pi} k h a \pi \omega \left( 1 - \frac{\pi \mu Na}{4r} \right) + P_e $$

(2.3.6-1)

As mentioned in sections 2.1 and 2.3.5, the nonlinear electric dissipation and the power consumed by auxiliary devices can be assumed as constants. So, adding them together into a constant, $H$, the above equation can be simplified into:

$$ P = R I^2 + H + \sum_{i=1}^3 \left[ B_i + E_i \left( \frac{N}{\pi} k h a \pi \right) + G_i \{ Na \} \right] \omega^i + \frac{N}{\pi} k h a \pi \omega \left( 1 - \frac{\pi \mu Na}{4r} \right) $$

(2.3.6-2)

Moreover, the last term on the right hand of the above equation can be embedded into the previous term, just by adding one unit to $E_1$, i.e. replacing $E_1$ by a new $E_1'$:

$$ E_{1,New} = E_{1,Old} + \left( 1 - \frac{\pi \mu Na}{4r} \right) $$

(2.3.6-3)

Consequently, the above equation can be simplified further into:

$$ P = R I^2 + A + \sum_{i=1}^3 \left[ B_i + E_i \left( \frac{N}{\pi} k h a \pi \right) + G_i \{ Na \} \right] \omega^i $$

(2.3.6-4)

$G_i$ can be embedded into $E_1$ and, therefore, set to zero. However, $G_2$ and $G_3$ are non-zero.

Equation (2.3.6-3) is the final form of the power consumed by a milling machine while cutting a slot of constant depth out of a workpiece. In this equation:

- $P$: Total power consumed by the machine
- $R$: The total apparent resistance of the machine’s linear circuits, viewed from the points that feed these circuits (To be measured)
- $I$: The fraction of the electricity entering the machine from the grid, which goes into the linear circuits of the machine
- $H$: Constant, depending on the nature of the machine’s non-linear electrical circuits and its auxiliary mechanical elements (To be found throughout the experiments on every single machine)
- $B_i, E_i$ and $G_i$: Constants, depending on machine’s geometry and the nature of its moving elements’ contact surfaces, its bearings and aerodynamics of its moving elements (To be found throughout the experiments on every single machine)
- $N$: Number of teeth on the milling tool
- $k$: A constant depending on the cutting process mechanism. $k$ is likely to change with changes in process parameters like speed, feed, depth of cut, angle of the teeth with the direction of cutting and the material itself. However, we assume that the change due to alterations in feed, speed and the depth of cut is negligible for a specific tool and material. The extent of the accuracy of this assumption must be verified through experiments.
The depth of cut
feed per tooth
the material’s shear strength
tool radius
spindle speed

3. CONCLUSIONS AND FURTHER WORK

Equation (2.3.6-4) gives the total power consumed by a milling machine while cutting a straight slot of constant depth out of a metal workpiece. After finding the 10 constants in this equation- \( R \), \( H \), \( B \), \( E \), and \( G \), it can be used for prediction of the power consumed by the machine during doing such a process with different feeds, speeds and depths of cut. This prediction of the total power consumption makes one able to design the process plan in a way to minimise to that or, more realistically, integrate the predicted consumed energy into the cost function of the process which has to be minimised during the process planning.

Since the terms in equation (2.3.6-4) come from different origins, after finding the constants, one can have an idea of the fraction of power used by different sources of consumption in a range of process parameters. This, not only helps the process planner to avoid the regions of cutting parameters in which the efficiency of the cutting process is critically low, but also enables a prospective user of machine tools to compare different machines from the energy consumption viewpoint.

However, it is required for the current \( I \) to be related to the total power consumed, \( P \) to make the model complete. A simple assumption of proportionality between current and power meaning a constant phase angle between electric potential and current vectors on the complex plane is a reasonable assumption to be made and examined. To be able to predict the total energy to be consumed for cutting workpieces in general, this model must be extended to include other cutting processes and the machine energy consumption during operations other than cutting, i.e. changing tools, taking measurements, etc.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by a number of grants for Engineering and Physical Sciences Research Council (EPSRC), involving a large number of industrial collaborators. In particular, current research is being undertaken as part of the EPSRC Innovative design and Manufacturing Research Centre at the University of Bath (reference GR/R67507/01).

REFERENCES

Role based Self-adaptation of a robot DiMS based on system intelligence approach

Mr. Kai Salminen, Mr. Hasse Nylund, Mr. Paul Andersson
1,2,3 Department of Production Engineering,
Tampere University of Technology
P.O BOX 589, 33101 Tampere, Finland
Tel: +358 3 31153295

ABSTRACT

Distributed Manufacturing Systems (DiMS) express self-adaptation as respond to changing service request over system life-cycle. For this purpose the roles and related changing expertise needs to be communicated. As a DiMS type self-integrating robot-holarchy can be seen as a fractal system without boundaries its structure and operations as well as combined expertise can be understood only through role based interactions and connections. The connections provide means and the interactions provide the combined systemic intelligence needed for actions. This paper studies the design of an adaptable robot holarchy. The influence of internal connections and the topology of network are significant. The negotiation connections are multiple for “stronger roles” and few for “weaker roles”. The holons representing stronger links form the core of a systemic “small world” and combine internally higher level services from the resources within the holarchy using internal logic frame based negotiation. Thus a fractal negotiation process is formed. The connection of the needed decision making and systems intelligence is done using a logical frame system. This allows the self-adaptation for common goals, for example the most sustainable solution. The approach is supported by a digital manufacturing approach allowing the comparison of real and virtual manufacturing data for learning. This research is a part of the test bed project for a holonic self adaptive manufacturing system.

1. INTRODUCTION

Self adapting manufacturing systems like Distributed Manufacturing Systems (DiMS) [1] are autonomous holonic systems on the ICT platform where communication is based on the use of Web Semantics and Service Oriented architecture (SOA. The self adaptation is based on efficient the re-use of knowledge gained from process and in process. As the field of change is multidisciplinary the adaptation process also needs to be capable of producing a good synthesis of requirements. This paper studies through a practical industrial case the design of a novel robot based production system that expresses improved self-adaptation ability. The proposed system consists of autonomous robot holons that form a modular holarchy. The adaptation itself is based on a systems intelligence build on knowledge embedded in Semantic Web Services that allow efficient digital manufacturing support. Adaptation is thus done virtually and test manufactured against digital models. The same model is used for dynamic production control, proactive quality check and learning from and within process.

2. SYSTEMS INTELLIGENCE BASED ADAPTATION

Efficient re-use of knowledge and related equipment is the prime prerequisite of a self adapting system. Only the availability of reliable services makes possible the rapid ramp-up of new cases. In dynamic world this services need to base on defined hypothesis and be tested and improved constantly. At the heart of the process lies meta-competence or new learning that can be transferred to new situations, thus forming the basis for self-organization abilities [2]. This kind of system allows the constant innovation process as each operation allows the testing of underlying hypothesis. Thus the system improves constantly the its ability to adapt to new requirements as it learns to use its capabilities and knowledge better and can create rapidly well defined new knowledge. Operational processes within system can then be autonomous, i.e., there is no need for online intervention, assistance or direct control of the operators by the rest of the organization. This principle has strong focus on operational processes,
using autonomous holons and holarchies, (man- machine teams), defining system measures and assessments to stimulate performance and developing the skills and capacities of system entities. Formalization of standard operational procedures to define the process and its control are the main means of governing the process and learning (on the concept of learning bureaucracy) [3]. These “best practices” are created through innovative synthesizing expert process as part of team work. The chosen methodology involves the workforce and form the basis of teaching and unifying the development of skills and competencies. This knowledge based approach supports innovation and gives good basis for coping uncertainty of the operational conditions and the nature of the market, the workforce and industrial relations, etc

2.1 ROLE BASED MODULAR SYSTEMS INTELLIGENCE AS BASIS FOR SELF ADAPTATION

A self organizing system grows organically however controlled to maintain the ability to integrate new members to existing structure. System emergence is guided to prefer similar or generally known cases that allow use of learning curve. Locally each joining entity (node) has to have a good knowledge level over the structure of neighboring nodes and also over both their current activities and coming activities. Broker holons that organize needed connections have an important role as they have enough links to gather the needed information Their role is on the one hand links service providers to service requestors and on the other hand makes self adaptation possible. Different System’s thinking views on holarchy type systems are usually limited to study different entities their properties and activities in relation to others. [4] [5]. Lately more effort is put on studies on the role of the internal negotiation process and the properties of negotiating network and its topology [6].

2.2 LEAN-META-MODELLING APPROACH AND LOGICAL FRAME SYSTEM FOR COMMUNICATING DESIGN INTENT

The Adaptation process needs good guidelines to control proper performance of emerging manufacturing solution. Adaptation system using lean-type meta model rules organizes naturally emphasizing short distances and minimal amount of links and activities. Also coherence checks against logical level system is vital as each activity represents additional cost, time, scheduling and quality issues as well as availability issues of for example transport, storage, tooling, fixturing etc.. Therefore the self adaptation process needs corresponding information to support organization decisions. Typical Lean-rules are for instance [2]:

1. Every operation is formally specified as to content, sequence, timing, and outcome. This is the central perquisite of sound exact quality. All the variation translates into poorer quality, lower productivity, and higher costs. More important, it hinders learning and improvement in the organization because the variations hide the link between how the work is done and the results.
2. Every connection must be standardized and direct, unambiguously specifying the unit involved the form and quantity of the goods and services to be provided, the way requests are made by each link, and the expected time in which the requests will be met. This rule creates a supplier-customer relationship between each entity that is responsible for providing each specific good or service.
3. Every product and service flows along a simple, specified path. That path should not change unless the adapted production line is expressly redesigned.
4. Rule for constant improvement and radical improvements. Any improvement to production activities, to connections between workers or machines, or to pathways must be made in accordance with the scientific method, under the digital manufacturing system.

These rules can be described within logical level system and addressed within digital manufacturing system using them in the concepting phase. Thus the models created out of concepts inherit the mind-set of the company and the known best-practice rules. For automatic model generation the rules have to be formally coded. The first rule requires the use of specific manufacturing receipt that covers the needed functionality. A manufacturing receipt represents operation and uses a logical level system in its representation. Thus it is possible to run the system using different meta- information. For instance the sustainable mind-set requires the selection of processes using least energy and other resources whereas an economically weighted process stresses machine utilization and efficiency.

A framework framework of neurological levels was developed by Robert Dilts [7]. Dilts attributed this system to the anthropologist Gregory Bateson, who introduced the notion of logical levels [8]. It was then adapted for Business Process Re-engineering and change management. The highest level in the logical level system is the vision level, which gives the purpose of the system. The second level is the strategic level, defining the goal, thus the emergence can be goal oriented. The third level is the value and paradigm level guiding the system emergence. Within this case is used lean-principles and values addressing sustainability, competitiveness and high added value. The fourth level is the identity level, on which the system gets its identity and role within whole and all the system entities their role and identity within the system. The role based approach uses this level for virtual adaptation. Each modular unit or
node within the system has, thus, the fifth level as the resource and skills level, describing the resources and skills available for executing the task. The sixth level describes the management of the autonomous unit – it relates, for example, scheduling, maintenance, and co-operation with other entities within system – and finally the seventh level describes the physical part and context. Each level represents a different conceptual negotiation process with its related ontology. The system supports the decision-making process by allowing internal coherence verifications. Within goal-oriented self-adapting systems changes on a higher level can be used to initiate corresponding changes on lower levels in a fractal manner. Change demands on a lower level can also challenge higher-level logic. By addressing test changes to a virtual system their effects on coherence can be evaluated.

![Logical frame system, use of belief system and values guide innovation](image)

The emergent adaptation has to be coherent and it should express the need whether it is the fastest alternative, most cost effective alternative or most sustainable alternative. Usually it is the local system within individual production specialized on specific services that has experienced a long learning curve and can therefore provide a complete and coherent total solution. Adaptation is therefore concentrated on improvements, innovations, capacity adjustments, new customer requirements rather than flexible routing, machine breakdowns or complete unknown new operations.

3. **Strong links as centres of self-configuration**

Within this kind of knowledge and skill based innovative self adaptation process the importance of system entities whose role is to act as strong links grows and the effective actions rely on them. If this kind of link is removed or renewed the affects for the network are considerable. If for instance the replacement process of a key resource does not take on account the continuum in system intelligence the results may be deteriorated ability to act on the network. Also if one of the strong knots suddenly starts acting against the interests of the network the consequences are severe [10]. Also as new resources are added their integration is not only an ICT-issue. There needs to be a system that provides the needed integration process. As the message broker accepts a resource to the system there has to be an integration negotiation.
The integration negotiation is dependent on the role of the unit in the system. Strong links have to negotiate with several other units and weak ones only with neighbouring units. As the system is fraktal only strong links need to communicate with other layers as similar data architecture allows the information to be spread to the whole system. Thus the information flow can be limited.

4. Systems Intelligence Based Emergency for Realizing Design Intent

The basic requirement for a self-adaptation process is that the system can achieve its purpose within changing dynamic circumstances. As discussed earlier the perquisites are a sound systematic for efficient re-use of knowledge, logical coherence and a good synthesis driven concept based digital manufacturing environment. The concept creation process is guided by system intelligence. System intelligence is semantically coded to a logical level system providing a sound conceptual framework. It is of vital importance that the right conceptual level can be addressed when a problem is encountered. As the system is designed its purpose is described within strategic level, the design guidelines within value and paradigm level, individual roles on identity level, usable resources and skills on resource level and management principles on management level. Thus it is possible to address each problem from higher conceptual level i.e., adaptation to machine failure is addressed from the management system whereas a resource problem is addressed from a strategic level. Within this research we used Lean-paradigm and corresponding design rules and methods derived from it as a test set of known best practices.
The Lean concept creation process has to express three basic Lean capabilities:

- Capability to integrate requirements from business, design and manufacturing domains for coherent design intent, as within today’s complex world the system has multiple objectives.
- Capability to constantly develop and learn, key elements in constant development are innovative co-operational processes consisting of continuous improvement based on rigorous testing against hypothesis.
- Capability to innovate and re-focus. The ability to pursue more than one objective is created by synthesizing towards radical innovation, which is necessary in order to face novelty.

To be able to fulfill these needs the concept creation process needs to be based on efficient synthesis methods and tools. The outcome has to be formalized to give basic guidelines for model creation. In order to efficiently use existing models and knowledge the similarity based reasoning support is needed. Semantic coding and use of related taxonomies and ontologies to support model creation is envisioned and under development.

5. **Intelligent Semantic Web Service as Means of Adaptation**

Individual autonomous units (holons) form an open complex system platform that can be also understood as a system theoretical “small world” a dynamic system. This system builds on top of the IP-network functioning next generation of the World Wide Web WEB 2.0. The core of a self-adapting manufacturing system is in its role based service negotiation process. A service is a standard software interface that has a knowledge part a needed software application and/or program with functionality of the process. A Semantic Web Service (SWS) is used as the basic value offerings given by a system as implementation of the service using web-based standards for the communication function. These service request or a service offerings contain semantic content that allows them to be found and used. In order to be used automatically this content should be formally described for a Semantic Web Service that enriches the service description with machine interpretable semantics. The role of services within DiMS is central. It represents the value offering of a manufacturing unit and, at the same time, it has the vital manufacturing knowledge embedded in it. As the logical level representation allows meta data, the underlying hypothesis can be tested by digital manufacturing.

6. **Design of an Adaptive Robot Based Manufacturing System in Industrial Context**

The approach discussed in this paper is studied within an Industry context. The given case was to create an innovative new production line for new innovative product line idea. The first stage was to test the developed integrated synthesis process with design, business and manufacturing domains using innovation tools and developed approaches. The outcome of the process was fed to logical frame system to ensure the coherence and give the system to be planned the needed reference points. Thus the system was given vision, goals and values. As for tests manufacturing paradigms to be used were chosen the first four layers of logical frame model were ready for use. From the manufacturing system planning point of view the most important guidelines are the definitions of product family architectures and platforms, quality concept, demand-supply concept and sales concept. The basic manufacturing paradigm used within the case company is Lean and the main principles of it were used in the system design. For manufacturing system concept creation were three alternative production automation paradigms chosen and tested by modeling and simulation against requirements. These were: [1]

- FMS (Flexible Manufacturing System) paradigm
- RMS (Reconfigurable Manufacturing system) paradigm and
- DiMS paradigm

The concept chosen for final modeling and testing was thus the one created on Lean-paradigm using DiMS approach allowing single piece flow and customer specific orders on rapid innovation oriented re-focusing of the system.

Three alternative customer specific value based modes were noticed, one for prize oriented, one for delivery time oriented and one for sustainability oriented. The basic difference for production modes is in the way a process is utilized. For prize oriented customers the main driver is costs, for delivery time oriented it is through put time and for sustainability oriented energy consumption and use of substances. Also the cost models and information needs are then different. For example the last group needs carbon footprint calculations.
The solution effectively used existing knowledge and experiences. The transition from concept to model required the modeling of intended robots, tools and grippers. Used methods and related manufacturing receipts were verified in laboratory environment and the robots actual capability tested by measuring and test manufacturing. Thus related error matrices were created for virtual model based simulation. Each tested manufacturing receipt corresponding to needed operation is now ready for formal description for needed web services (SWS) . The next phase of the project is to formalize the logical level based system intelligence solution to support needed adaptability during operation. The needed taxonomy and related ontology structure with supporting semantics is currently under development. This is to be accomplished June 2009.
7. Conclusions

The design of a knowledge-based adaptable system capable of innovating itself requires a thorough synthesis between manufacturing business and design domains. To enable efficient adaptation, the introduced holonic structure offers adequate modularity to support constant innovation and improvement of individual units without affecting the whole. From a systems intelligence point of view, the structure has to be composed of leading holons embedded with core competences linked efficiently and weak holons performing simple tasks needing only single links. Use of logical levels and related reasoning requires still a lot of semantic work and building improved ontologies. The used approach allows holon-based system intelligence modularity on holonic base where the individual holons can communicate their knowledge with simple SWS negotiation. The needed adaptation work reduces thus significantly. The whole system is still to be tested and verified within real manufacturing but already at this stage the approach has proven its possibilities. The created solution is already more flexible, efficient, productive, sustainable than compared traditional FMS and allows new features to product compared to solution used for previous product line.

REFERENCES


Effects of Chemical Functionalized MWCNTs/Epoxy Matrix Composites on Electrical Conductivity Behavior

Abu Bakar Sulong\textsuperscript{1}, Norhamidi Muhamad\textsuperscript{1}, Che Husna Azahari\textsuperscript{1}, Ahmad Rasdan Ismail\textsuperscript{1}, Joohyuk Park\textsuperscript{2}

\textsuperscript{1}Dept. of Mechanical and Materials Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia

\textsuperscript{2}Dept. of Mechanical Engineering, School of Engineering, Sejong University, 98 Gunja dong, Gwanjin gu, 143-747 Seoul, South Korea

E-mail: jhpark@sejong.ac.kr
Tel: +822-3408-3771; Fax: +822-3408-3333

\textbf{Abstract}

Few attempt to study the effects of chemical functionalized carbon nanotubes (CNTs) on the electrical conductivity of nanocomposites, instead to increase dispersion and interfacial bonding strength between CNTs and polymer matrix for improvement of mechanical properties. Therefore, in this study the electrical conductivity of two types’ chemical functionalized (Carboxylated and Octadecylated) multi-walled carbon nanotubes (MWCNTs), non ionic surfactant additive MWCNTs and as produced MWCNTs epoxy nanocomposites are investigated as function of MWCNTs concentration. As produced MWCNTs and surfactant additive MWCNTs nanocomposites gave higher electrical conductivity than functionalized MWCNTs nanocomposites. However, chemical functionalization of MWCNTs significantly decreased the electrical conductivity of nanocomposites. This behaviour is further investigated through image analysis of powder CNTs and formation electrical conductive pathway network under applied electric field. It can be deduced that the chemically functionalized MWCNTs through severe acidic treatment are not suitable for electrical applications.

\section*{1. INTRODUCTION}

Single walled carbon nanotubes (SWCNTs) consisting of a single graphite plane (so-called graphene) rolled up into a cylinder behave as either a metal or a semiconductor depending on their wrapping angle of graphene sheet and its diameter, while MWCNTs are reported to be always electrically conductive and to have an electrical conductivity approximately $1.85 \times 10^3$ S/cm [1-4]. CNTs polymeric composites are aimed at the exploitation of the high electrical conductivity of CNTs coupled to high mechanical properties, thermal properties and others unique properties [5-7]. Due to the low filler loading fractions required, the mechanical properties and the surface finish of the composite matrix can be maintained. In contrast, conventional surface coating such as Indium Tin Oxide (ITO) coating offer a high electrical conductivity but very high production cost and low quality of surface finish. Therefore, alternate approach using CNTs as conductive fillers in polymeric matrix promises a unique combination of low cost production, ease processing with high electrical conductivity. Indeed, these materials are increasingly used in the electronics sector, as well as for automotive and aerospace applications, where an electrical conductivity exceeding $10^{-8}$ S/cm is required for dissipation of electrostatic charges.

Recently, many researchers have been interested in chemical functionalization of CNTs to achieve the better dispersion of CNTs in polymer matrix and increase the interfacial bonding strength between CNTs and polymer matrix [8-13]. From our previous study, incorporated chemical functionalized MWCNTs improve dispersion quality...
MWCNTs in epoxy matrix and increase the tribology properties of epoxy nanocomposites than As produced MWCNTs epoxy nanocomposites [14,15]. Moreover, Scanning Electron Microscopic (SEM) image analysis of fracture surfaces indicated that chemically functionalized MWCNTs have higher interfacial bonding strength with epoxy matrix than As produced MWCNTs. A high electrical conductivity at low loading concentration of As produced SWCNTs and As produced MWCNTs have been reported [16-19]. However, there were few attentions to study the effects of the chemical functionalization of CNTs on the electrical conductivity of epoxy nanocomposites. Therefore, the electrical conductivity of two types chemical functionalization (Carboxylated and Octadecylated MWCNTs), non ionic surfactant additive MWCNTs and As produced MWCNTs epoxy nanocomposites are investigated in variation of CNTs loading concentration.

2. EXPERIMENTAL

The epoxy resin is a bisphenol-A-based epoxy resin (KER 215), which contains mono-epoxidized alcohol as a reactive diluent (Kumho P&B Chemical). The non-MDA aromatic amine curing agent (Amicure 101) is supplied by Daemyung Chemical Tech. The viscosity of the epoxy resin is reported as 0.7~1.1 Pa s and the curing agent as 0.2 Pa s. A lower viscosity of epoxy matrix is selected for better wetting conditions with reinforcement fillers. The reinforcement filler used in this study is MWCNTs. It has been synthesized by thermo-chemical vapor decomposition of hydrocarbon gases. Two types of functionalized MWCNTs were used, Carboxylated MWCNTs (MWCNTs-COOH) and Octadecylated MWCNTs (MWCNTs-CONHCH₃(CH₂)₁₇) [8,14]. Characterization of As produced MWCNTs and chemical functionalized MWCNTs were conducted by Thermogravimetric (TGA), Fourier Transform-infrared Spectroscopy (FT-IR), Raman Spectrometer and Transmission Electron Microscopic (TEM) [14]. Moreover, the optimization of dispersion time and process had been also discussed elsewhere [14]. The dispersed MWCNTs in epoxy mixture was injected to Teflon mold and cured under pressure with the hot press to minimize existing voids in epoxy nanocomposites. Finally, the bulk MWCNTs epoxy nanocomposites were machined to specimen with dimension of 13 mm (width) x 25 mm (length) x 2 mm (thickness) for the electrical conductivity measurement as shown in Figure 1. The measured surfaces area of specimen as given in Figure 1 were polished with sand paper, and were pasted with uniform thin thickness of high conductive silver paste to minimize contact resistance during measurement. The electrical measurement system used in this study is shown in Figure 2.

Figure 1: Specimen for electrical conductivity (a) Pure epoxy polymer, and (b) Epoxy Matrix with MWCNTs (note: upper and bottom surface pasted with silver paste)

Figure 2: Electrical conductivity measurement system specimen for electrical conductivity
3. RESULTS AND DISCUSSION

Figure 3 is representative obtained results from electrical measurement system given in Figure 2. Figure 3 shown curves of the variation of current intensity as function of voltage. Linear fit for resulted curves had been applied to obtained constant tangent value of incline, represent as resistance (R) which follow Ohm’s Law [16].

Figure 3: Representative of variation Current Intensity as function of Voltage results and it’s linear fit lines

The variation of electrical conductivity of epoxy nanocomposites as function of MWCNTs types and loading concentration in weight percentage (wt%) are plotted in Figure 4. Non-conductor material of pure epoxy polymer become conductive as incorporated CNTs. At low loading concentration As produced MWCNTs gave a high electrical conductivity, which similar results were reported elsewhere [16-19]. The electrical conductivity of epoxy nanocomposites are increased with increasing of CNTs loading concentration. In our previous study reported that chemical functionalized MWCNTs improve the dispersion quality and strengthen of interfacial bonding strength with polymer matrix [14,15]. However in this study, incorporated chemical functionalized MWCNTs in epoxy matrix resulted in significantly decreasing of the electrical conductivity of bulk epoxy nanocomposite. For dissipation of electrostatic charge, it is necessary to load more than 1.0 wt% concentration MWCNTs of surfactant additive MWCNTs, As produced MWCNTs and Octadecylated MWCNTs. However, for Carboxylated MWCNTs, it required loading more than 3.0 wt% for dissipation of electrostatic charge. Fractured surface SEM images of As produced and Carboxylated MWCNTs epoxy nanocomposite are given in Figure 5. Even though there are fiber pull-out phenomenon on fractured surface of As produced MWCNTs epoxy nanocomposite, dispersion level for both types of CNTs is in similar manner. Significant CNTs agglomeration also cannot be observed on both types of CNTs, which may play role as electrical conductive pathway in polymer matrix. Therefore, influences of chemical functionalized CNTs on the electrical conductivity behavior are further investigated. For visualize this behavior, an electric field was applied to 0.01wt% of As produced MWCNTs and Carboxylated MWCNTs in ionic alcohol (Isopropyl Alcohol) as electrolyte through 1 mm distance between of two Cupper electrodes, respectively. The movements of CNTs under given applied electric field were observed by an optical microscope.
Figure 4: Variation of electrical conductivity of epoxy nanocomposites as function of MWCNTs loading concentration and MWCNTs types.

Figure 5: SEM fractured surface images of (a) As produced, and (b) Carboxylated MWCNTs epoxy nanocomposites

Theoretical consideration states that in the presence of an electric field $\vec{E}$, each conductive nanotube experiences a polarization $\vec{P}$. This polarization can be divided into two contributing component, i.e. one parallel to the tube axis $\vec{P}_\parallel$ and one in radial direction $\vec{P}_\perp$. The magnitudes of both components depend on the polarizability tensor of the nanotube. For single-wall carbon nanotubes, it has been already been suggested that the static polarizability in the direction of the tube axis is much larger than across the diameter [20]. This polarization leads to a torque $\vec{N}_E$ acting on the nanotube. Under given condition, this torque aligns the nanotube against the viscous drag of the surrounding
medium in the direction of the electric field. An illustration of the behavior of a cylindrical particle exposed to a homogeneous electric field is given in Figure 6.

Figure 6: Schematic illustration of a polarized cylindrical particle in an electric field.

During the application of a direct current (DC) electric field, a fraction of the CNTs are observed to move forward to anode, under electrophoresis, verifying the presence of negative surface charges. As soon as these CNTs are close enough to the electrode to allow charge transfer, the CNT is discharged and adsorbed onto anode. Tips of CNTs are connected to the electrode then become sources of very high field strength and the location for adsorption of further fillers particles. As a result, ramified CNTs network structures extend away from the anode, eventually reaching the cathode and providing electrical conductive pathways throughout the sample. Above theoretical behavior had been observed at As produced MWCNTs under electrical field as given in Figure 7. A comparable behavior also has been observed for MWCNTs in different solvents [21,22].

Figure 7: Electrical conductivity behavior of As produced MWCNTs in Isopropyl alcohol while applied (a) 200V/cm, (b) 400V/cm, and (c) 600V/cm electrical field.

The results of CNTs electrical conductive pathways formation while increasing electric field intensity from 200 V/cm to 600 V/cm are given in Figure 8. However, CNTs network structures which work as electrical conductive pathways are not occurred at Carboxylated MWCNTs. It clearly can see that most of Carboxylated CNTs are
attached to the anode electrode, where thicker CNTs agglomerations are observed at anode electrode rather than cathode electrode. It assumes that carboxylic groups give unbalance polarization on CNTs walls, where it contributes more negative ion on CNTs surface. This lead most of the Carboxylated CNTs have more tendencies attracted and attached to the anode rather than formation of electrical conductive pathways as observed in Figure 7. Without formation electrical conductive pathways, it wills results on no electric current transfer between electrodes. During introduction of carboxylic group (COOH-) through chemical reaction by refluxing inside concentrated acidic solution may cause the physical structure damage to the wall of CNTs. These carboxylic groups are then chemically reacted with octadecyl-amine to form Octadecylated MWCNTs, which able to reduce the polarization effect by resulted in slightly improvement of the electrical conductivity bulk epoxy nanocomposite. Non-ionic surfactant was used as dispersion agent for As produced CNTs give the best electrical conductivity at low concentration than other types of CNTs. Only surfactant additives MWCNTs epoxy nanocomposite become conductor at 0.1 wt%. This behavior is not influence from the surfactant properties because TRITON X-100 which used in this study, is non ionic surfactant where it can’t work as electrolyte for allow electric current flow through it. It assumes that by addition of surfactant to the epoxy matrix give better wetting condition with CNTs than As produced CNTs. Finally, powder As produced and Carboxylated MWCNTs were observed by SEM as given in Figure 9. Figure 9 (b) shown that after chemical functionalization through acidic treatment, length of individual CNTs have been shortened. It assumes that physical structure of As produced CNTs had been damages, where these damages may interrupt the chirality of CNTs. The chirality of the CNT has significant implications on its material properties. It has a strong impact on the electrical properties of the CNT. Graphite is considered to be semi-metal, but it has been shown that CNT can be either metallic or semi conducting, depending in the tube chirality. All the armchair type CNTs are metallic, whilst this is only in the case for chiral or zigzag CNTs if \((n-m)/3\) is whole integer, otherwise, they are semi conducting [2]. It assumes that longer tubes of As produced CNTs in Figure 9 (a) give more advantages in the formation of electrical conductive pathways, where it ease to connected between individual CNT.

Figure 8: Electrical conductivity behavior of Carboxylated MWCNTs in Isopropyl alcohol while applied (a) 200V/cm, (b) 400V/cm, and (c) 600V/cm
4. CONCLUSION

This study’s results indicated that chemical functionalized CNTs significantly decreased the electrical conductivity of epoxy nanocomposites due to unbalance polarization effect and physical structure defects due to severe condition during acidic treatment process. However, As produced CNTs and As produced CNTs with additional surfactant give the higher electric conductivity of epoxy nanocomposites than incorporated with chemical functionalized CNTs. Therefore it can deduce that non chemical functionalized CNTs are more suitable for the electrical applications. But, chemical functionalization of CNT is still necessary for increase dispersion quality and strengthens the interfacial bonding strength with polymer matrix, which more important in structural applications.

ACKNOWLEDGEMENTS

This work was supported by grant No. R01-2003-000-10-72-0 from the Basic Research Program of the Korea Science & Engineering Foundation. Sulong, A. B. acknowledges a fellowship support from Sejong University, Universiti Kebangsaan Malaysia and Ministry of Science, Technology and Innovation (Malaysia).

REFERENCES


The Feasibility of Printable Power

Huw J Lewis*

Manufacturing & Operations Engineering Department
University of Limerick
Limerick, Ireland

ABSTRACT

The use of printing techniques as a manufacturing process is well established, in the form of 3D printing rapid prototyping systems. Additive methods of manufacturing of this form are being considered as environmentally friendly as less material is wasted and the number of processes needed for the finished product is reduced.

Printed manufacturing is being established in the electronics industries as conductive inks and electrical conductive adhesives have been developed. This has allowed for the creation of printed circuit boards, on various substrates using screen printing methods and developments towards the use of inkjet printing for circuit design. Coupled with this, a range of printing methods can now be used to produce components such as transistors, diodes, sensors and RFID tags. Whilst printed batteries have been developed they tend to be physically large to allow for a suitable discharge life.

This paper investigates the feasibility of screen printing a silver-zinc battery, using off the shelf conductive inks, and developing a screen printable electrolyte. It presents the development of the cell shape, and gives a comparison of the electrolyte strengths. The voltage discharge curves for the cells are shown, with and without a printable separator medium, and the stackability of the cells are tested.

Whilst initial tests of the battery are promising further development of the cells are needed to increase their efficiency. This will include the development of the separator material, and cell shape, before manufacturing the required screens.

1. INTRODUCTION

The use of printing techniques on a range of substrates [1] is being used in the electronics industries for the production of circuits and individual components such as transistors, diodes, sensors and Radio-frequency identification (RFID) tags [2]. Well established printing methods such as flexography, gravure, pad, offset lithography, screen and more recently ink-jet printing are being utilized in the industry, due to their ability to print on flexible substrates, the low costs involved and the high speeds and volumes attainable in the process. The process builds up a circuit or component by adding layers of different materials where they are needed, thus reducing waste. These techniques are being considered by a number of companies to develop power systems in the form of batteries [3]

1.1 Electrical Cells.

The creation of electrical batteries or cells can be attributed to Volta in 1800. The principle of the battery is based on the chemical reactions of oxidation and reduction. [4]

“Oxidation is the process of releasing electrons whilst reduction consumes electrons”

Thus in order to develop an electrical power source it is necessary to have an electron source that will oxidise (anode), a receiving source to accept the electrons (cathode), an electrolyte that will allow the movement of ions, and a separator that ensures that the electrons flow through to an external circuit when connected. (Fig1).

In this case it is intended to utilize the screen printing process and standard off the shelf silver and zinc screen printing pastes [5] as the anode and cathode, fixing the size of the both these elements and varying the electrolyte. The
performance of the battery without a separator will be analysed as well as the possibility of using a non corrosive electrolyte.

1.3 Screen Printing.

Screen-printing uses a printing screen or stencil which consists of a frame with a stretched a fabric mesh. The mesh has open apertures which form the desired pattern to be printed. As the ink or coating is forced through the open design (by a squeegee), the pattern on the screen is replicated on the substrate [6] (Fig 2).

The screen-printing process is essentially a simple efficient method of reproducing patterns on a variety of substrates, which include plastics, fabrics, metals and papers. When screen-printing, there are a number of key operating parameters of which screen type, (mesh and emulsion thickness) conductive ink properties (viscosity, particle size and solids content) are among the most important. However the key to screen-printing reliable electronic circuits depends on monitoring the printing parameters and maintaining specified limits. These limits determine if the print will be successful and indeed which process parameter settings offer the best opportunity to print error-free. These parameters will be investigated once a suitable cell design has been found.

2. Experimentation

The initial aim of the analysis was to investigate whether an electrical potential could be developed utilising screen printable inks (Table 1), and a range of electrolytes (Table 2). Tests were performed by immersing separate plates of printed silver and zinc pastes into an electrolyte, and measuring the voltage output. The plates (25mm x 25mm) (Fig 3) were printed to a thickness of 0.8mm (post curing) on either the copper side of standard copper clad printed circuit boards or on the epoxy surface where copper was removed. The distance between the plates was fixed using a 5mm diameter glass rod (Fig 4), with a data logger providing a load of 60kΩ (slightly less than the 6.2 multiple recommended by ANSI [7]). In both cases a separator was not used.

Table 1: Pastes used to produce a Printable Power Source

<table>
<thead>
<tr>
<th>Paste</th>
<th>Code</th>
<th>Fineness of grind</th>
<th>Metal content</th>
<th>Viscosity at 25°C</th>
<th>EN10204 Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3rd streak</td>
<td>Main Break</td>
<td>Shear rate - 1</td>
<td>Viscosity Pa s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silver</td>
<td>C2040922D1</td>
<td>4μm</td>
<td>1μm</td>
<td>62.89%</td>
<td>7.788</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.760</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.870</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.701</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.484</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinc</td>
<td>C2060918D5</td>
<td>6μm</td>
<td>2μm</td>
<td>80.26%</td>
<td>4.774</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.768</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.713</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.659</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.696</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.655</td>
</tr>
</tbody>
</table>
Table 2: Electrolytes Used in Producing a Printable Power Source

<table>
<thead>
<tr>
<th>Electrolyte</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Petri dish tests only</td>
</tr>
<tr>
<td>Sodium Hydroxide (NaOH)</td>
<td>Petri dish tests only 32%,20% [8]</td>
</tr>
<tr>
<td>Gelatin + water</td>
<td>at required viscosity</td>
</tr>
<tr>
<td>Gelatin + Sodium Hydroxide (NaOH)</td>
<td>at 12.5%,22%, and 30%</td>
</tr>
<tr>
<td>Gelatin + Sodium chloride</td>
<td>at 12.5%,22%, and 30%</td>
</tr>
</tbody>
</table>

The outcome of these initial tests can be seen in Fig 5, which show that the highest potential (approximately 1.05V) is developed using sodium hydroxide, with the pastes printed on the epoxy substrate. The electrical potentials formed in these tests are approximately 3.5 times greater than those where the pastes are printed on a copper substrate. This can be explained by the reaction between the copper and pastes Equation 1,2 [9], which was accompanied by the lifting of the pastes from the substrate (Fig 6). The electrons react between the zinc and copper, and the silver and copper, thus reducing the movement between the zinc and silver.

\[ \text{Zn} + \text{Cu}^{2+} = \text{Zn}^{2+} + \text{Cu} \]  
\[ \text{Ag} + \text{Cu}^{2+} = \text{Ag}^{2+} + \text{Cu} \]  

In the majority of other cases the substrate had little or no effect on the potential formed, with the potential being in the range of 0.5-0.6Volts.

![Fig 3 Silver and Zinc Test Pieces](Image)

![Fig 4 Test set up for Separate Plate Silver-Zinc Cell](Image)

![Fig 5. Initial Tests of printed plates with varying Electrolytes](Image)
Having established that the component parts of the cell will produce a potential, it is necessary to investigate the design of a cell that can be printed.

3. Printable Cells.

In the initial tests the cells were simply stacked, and immersed in the electrolyte. Attempting to produce a horizontal cell to match this would require the following sequence:

- Print and cure the bottom electrode (zinc) over a connecting track
- Print the electrolyte
- Print and cure the top electrode (silver) and connecting track

In order to print the different electrolytes, it was necessary to increase their viscosity; this was achieved by adding gelatin until the required constituency was obtained.

The difficulty in developing a vertically stacked cell is that the electrolyte has to remain viscous after printing, thus when attempting to print the top electrode, the electrolyte is either removed by the action of the squeegee, mixes with the electrode paste, or the electrode paste fails to adhere to the surface of the electrolyte. This causes problems in curing the electrode, in the form of surface cracking, and/or the evaporation of the electrolyte (Fig 7). Thus further experimentation was carried out using a horizontally stacked cell.

3.1 Horizontally Stacked Cells.

Horizontally stacked cells have the zinc and silver printed adjacent to each other with a small gap separating them (Fig 8). It is intended that the gap would accommodate the separator in future tests. In developing the suitability of such a layout, the size of the electrodes (zinc and silver) are fixed at 25 x 5mm, and the electrolytes have been
 predefined (Table 1), therefore, the variable within the process is the gap between the electrodes. The tests were carried out by immersing the cells into the electrolyte, but this will be printed in later tests.

The effect of varying the gap is shown in Fig 9. Whilst the 5mm gap produces a higher initial voltage when using NaOH (either 20 or 32%), it tends to fall away. As the outputs using H2O and NaCl seem to be higher using a 3mm gap, and the voltages for NaOH seem to stabilise quicker, it was decided that a 3mm gap would be utilised for further tests.

These initial tests indicate that a cell could be produced utilising zinc and silver pastes and overprinted using a relevant electrolyte. Whilst being able to form the cell, in a horizontal format two other problems exist, the printing of a separator, and the sealing of the cell, to reduce the drying out of the electrolyte.

3.2 Development of a Testable Horizontal Cell.

Having defined the initial parameters for a cell by immersing the plates into an electrolyte, it is necessary to produce a cell with a printable electrolyte. This was achieved by increasing the viscosity of the electrolytes (Table 1) by adding gelatin, and in the prototype stages, using a squeegee and mask to apply the electrolyte across the electrodes and the 3mm gap of the horizontal cell (Fig 8). To reduce the speed of evaporation, the electrolyte was sealed using a non conductive lacquer [10]. The tests were carried out without a separator between the electrodes, (i.e with the gap filled with the electrolyte), and with a porous separator filling the gap, in this case a printable mixture of talcum powder and water.

The cells produced (Table 3) were loaded with a 60KΩ resistive load supplied by the data logger for 1 hour in each 24, with the physical condition of each being analysed each day.

Fig 9 Voltage Output for Different Electrolytes with Varying Gaps between the Electrodes
Table 3. Parameters for Horizontal Cell Tests

| Electrodes  | Zinc: 25x25x0.8mm thick  
| Silver: 25x25x0.8mm thick  
| Gap between Electrodes: 3mm  |
| Electrolyte | Na OH: 22%  
| Na OH: 32%  
| Na Cl: 22%  
| Na Cl: 32%  |
| Sepatator | None  
| Talcum  |
| Sealant | Non- Conductive lacquer [10]  
| Substrate | Standard PCB epoxy board  |

The results (Fig 10) indicate the output voltages for the cells. It is apparent in both cases that the separator plays a major part in the performance of the cell, giving a higher initial voltage, and a shallower decline, albeit to a relatively low voltage. It is interesting to note that a higher output was gained with a weaker solution of NaOH, but a stronger solution of NaCl. In comparing the Outputs of the cells that have a separator, and either 22% NaOH, or 30% NaCl as the electrolyte (Fig11) it can be seen that the overall output from the cell with NaOH as its electrolyte is higher, the performance is more erratic over the test days, and the cell fails after 3 days. The NaCl electrolyte cell has a constant pattern of discharge as shown by the equations for the curves (Table 4), and a longer life pattern. It is felt that the reason behind the erratic performance of the NaOH electrolyte cell is the poor seal obtained using the insulating lacquer, the lack of ventilation during the discharge, and the drying and expanding of the separator (Fig12). In comparing these results with the initial tests (Fig 5), it is obvious that the NaOH electrolyte will perform at a higher level, if the separator and sealing can be modified, whilst the cells with NaCl are performing at a similar level as the initial results. In both cases the cells are performing below the possible 1.5V level that is deemed normal for silver zinc batteries. In order to increase the output using the current parameters, the stackability of the cells into batteries was investigated.

Fig 10 Voltage Outputs for Different Electrolytes with and without a Separator
3.3. Developing a Battery from Cells.

A battery is a combination of cells utilized to increase the desired voltage to a defined level. Two cells were connected, in order to increase the output voltage (Table 5). The ineffectiveness of the separator material is seen by the averaging of the voltage output between the two cells for the NaCl, and the 33% increase in output for NaOH. Further to this voltage is produced by connecting the zinc electrodes and measuring across the silver electrodes or vice versa, again indicating the problem with separator.

### Table 4. Discharge Equations for 33% NaCl Electrolyte

<table>
<thead>
<tr>
<th>Hour</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Hour</td>
<td>( V = 0.88 - 6.44 \times 10^{-6} t^2 + 22.82 \times 10^{-9} t^3 )</td>
</tr>
<tr>
<td>2nd Hour</td>
<td>( V = 0.69 - 5.76 \times 10^{-3} t^2 + 21.40 \times 10^{-6} t^2 - 28.10 \times 10^{-9} t^3 )</td>
</tr>
<tr>
<td>3rd Hour</td>
<td>( V = 0.64 - 5.50 \times 10^{-3} t^2 + 21.10 \times 10^{-6} t^2 - 28.60 \times 10^{-9} t^3 )</td>
</tr>
<tr>
<td>4th Hour</td>
<td>( V = 0.64 - 5.30 \times 10^{-3} t^2 + 19.10 \times 10^{-6} t^2 - 23.00 \times 10^{-9} t^3 )</td>
</tr>
</tbody>
</table>

\( V = \text{Voltage} \quad t = \text{time (seconds)} \)

### Table 5. Output Voltages for 2 Cell Combinations

<table>
<thead>
<tr>
<th>Electrolyte</th>
<th>Cell 1</th>
<th>Cell 2</th>
<th>Cell Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volts</td>
<td>Volts</td>
<td>AgZn - AgZn</td>
</tr>
<tr>
<td>30% NaCl</td>
<td>0.3</td>
<td>0.47</td>
<td>0.36</td>
</tr>
<tr>
<td>30% NaCl with Separator</td>
<td>0.39</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>22% NaOH</td>
<td>0.5</td>
<td>0.8</td>
<td>1.13</td>
</tr>
<tr>
<td>22% NaOH with Separator</td>
<td>0.95</td>
<td>0.78</td>
<td>1.17</td>
</tr>
</tbody>
</table>
4.0 DISCUSSION

The use of off the shelf materials and a standard screen printing unit has been investigated as a possible method of producing electric cells. The initial tests have shown that a cell can be produced by this method, but the efficiency and outputs are limited. It is likely that the cells that will be produced with the parameters stated in Table 3 will not be stackable as indicated in Section 3.3. Due to the mainly manual method of producing the cells the repeatability of the process could be questioned. Thus in order to further develop the feasibility of screen printing a cell, the shape and size of the electrodes together with the alternate methods of layering the cell need to be investigated. This will compliment the further investigation in to the separator materials and sealing methods, and the optimizing of the screen printing parameters. Alternate electrode materials will also be considered.

5.0 CONCLUSION AND FURTHER WORK

It was found that it is feasible to develop a cell with a limited voltage output (0.8-1.0 Volt) and life (up to 4 days) based on “off the shelf” printable silver and zinc pastes. In developing the cell as a horizontally stacked unit (i.e., silver and zinc pastes printed adjacent to each other), it was found that a 3mm gap was required for the separator material and electrolyte. The two electrolytes utilized in this paper NaCl, and NaOH, where diluted to 22% and 30%, with the lower percentage NaOH giving the better overall output and life. Testing of the cells for 1 hour in 24 indicated a fast discharge rate (e.g. 60% of initial output in 20 min on 1st test 22% NaOH), and a short life cycle (approximately 4 cycles /days). Part of the short life cycle may be attributed to the sealing of the system which showed degradation during and after the tests were completed.

Further development of the shape of the cell will be investigated, as will the use of a range of separator materials, electrolytes, and sealing methods, in order to increase the life and output of the cells. Coupled with this the optimum screen sizes and printing conditions need be formulated.

References.
ecoFLEX - Inter-branch Methodology for Flexibility Measurement of Production Systems

Sven Rogalski, Jan Siebel, Jivka Ovtcharova

ISPE/PDE
FZI Forschungszentrum Informatik
Karlsruhe, 76131, Germany

ABSTRACT

Demands on production systems are changing constantly as a result of changing competitive conditions and are linked to the performance goals of time, quality, cost and innovation ability. The increased uncertainty of planning, regarding kinds and volume of producible products, is an especially complicated challenge for production companies, which therefore leads to a growing demand of flexibility. In this context, methods for estimate the flexibility of production systems will be more important. The following article presents the construction and the methodology of an innovative evaluation technique called ecoFLEX, which allows production planners and manager to create reliable statements about their production systems. This paper details the general concept for estimating the flexibility of volume, mix and expansion, and the mathematical approach to calculate concrete ratios for these flexibilities on different levels of the production system - from the workplace level to the factory level.

1. INTRODUCTION

The basic conditions of economic production have changed greatly in the last years and have enormously increased the importance of flexibility of production systems. In particular, the handling of the continuously growing uncertainty of planning, regarding kinds (product/variant mix) and volume (output) of producible products, is a pivotal success factor for protection of competitiveness [1-2]. As a reaction to this, production companies are forced to adapt their systems, strategies and concepts to permit enough freedom to handle these uncertainties [2-4]. Presently, in intensively followed strategies, like Agile Manufacturing, Mutability and Holistic Production Systems, consider the flexibility as well as further associated properties of adaptation and modification as an important goal. They influence the different concepts about production system design and organization, which directly depend on production strategies [2-3]. In today’s turbulent operational environment, there exist many methods to realize flexible production: outsourcing, acquisition of highly automated manufacturing systems, insourcing and increasing of stock or transition to flexible working hours [2, 5-8]. Practice shows that these methods can contribute to more flexible production within a planned range [6, 9-10]. But the problem is that there does not exist satisfying technique to estimate the flexibility to all to evaluation of the lack of flexibility in production systems [4, 8-9, 11]. The absence of such bases of estimation, especially in product mix and necessary volume, is a tough challenge for production companies to find the optimal degree of flexibility in the application of such methods and guarantee an economical production in this way [1, 8, 11].

For example, out of fear of image and turnover losses, a company builds a high potential of flexibility to prevent having demands which cannot be met. But the potential of flexibility will not be used, leading to needlessly high additional costs that can be dangerous for a profitable production. Vice versa is also true when the potential of flexibility is lacking. This will lead to adverse effects caused by continuous, quick, unplanned and uncoordinated system adaptations. As a result, an economically worthwhile balance should be found between the uncertainties and the necessary degree of flexibility, as shown in the following figure.
2. CURRENT CHALLENGES FOR COMMAND OF FLEXIBILITY

The previous figure leads production companies to the question: how can we calculate the degrees of freedom and variety in production in reaction to fluctuations in volume demands? These kinds of uncertainties are not only relevant from an operational point of view, but also from a strategic perspective. Therefore, understanding the possibilities for capacitive expansions is also a key issue. Moreover, it is important to know how changes of the product mix affect the economical production. In this context, it should be possible to analyze the flexibility on different levels of view, because driver of production changes can influence different levels of a system. Hence an approved methodology for estimating the flexibility has to allow a uniform and objective evaluation for the following three kinds of flexibility:

- volume flexibility
- product/variant mix flexibility
- expansion flexibility

The volume flexibility is a measure for the degree of capability to react to changes in demands, especially to increasing demands. The mix flexibility is the grade of stability and freedom concerning the product/variant mix of production systems. Changes in demand and offer have not only operational importance, but also strategic relevance. In this context, expansion flexibility plays a prominent role in estimating opportunities to modify systems for the permanent increase of capacity [11-12]. In spite of the high importance and existing various solution methods, such methods are rarely used in industrial practice. In the literature there are many approaches. This shows there is great interest in this topic, but also the lack of a generally accepted method. Intensive inquiries for selective approaches showed there are many methods for estimating volume, mix or expansion flexibility. But, comparison of those methods shows that they follow different terminologies and are usually developed for a specific problem; the applications of the methods are restricted to their specific problems. Hence comprehensive and objective estimations regarding volume, mix and expansion flexibility on different levels in production systems are not possible, which shows the necessity for novel solution approaches.

3. THE ecoFLEX METHODOLOGY

The following section will introduce the concept of the ecoFLEX methodology, which was developed at the FZI Forschungszentrum Informatik in Karlsruhe/Germany.
3.1. DESIGN AND FUNCTIONALITY

ecoFLEX is an innovative evaluation methodology, which allows objective and inter-branch evaluations about volume, mix or expansion flexibility, based on concrete calculations. It integrates different calculation methods, called flexibility metrics. These consider in addition to aspects of times and costs also different levels of views, the workplace-, line-, segment- and factory view. Furthermore ecoFLEX ensures practicable usage of flexibility metrics for the production and change management. This is achieved using a production system model, a scheme of production infrastructure objects to be evaluated as well as their relations and dependencies. The connection of the flexibility metrics with this model is the precondition to analyze the adaptability of production facilities; furthermore, it allows for estimation of possible alternative configurations (cp. Figure 2).

ecoFLEX supports production planners and managers in making reliable evaluations about the flexibility of their production systems, and makes it possible to react to the increasing complexity and dynamic as well as decrease reaction time to service the market. As a result, ecoFLEX is an instrument for decision support, which can analyze and compare both existing production infrastructures and infrastructures that are still to be planned. In this context analyses between production systems from various branches are made possible. This allows analogous systems to reach novel conclusions that can be shared and lead to sustainable structures.
3.2. CONCEPT OF EVALUATION OF FLEXIBILITIES IN ecoFLEX

After the previous description of the current challenges for command of flexibility and the design of ecoFLEX, the following sections will explain how to evaluate volume, mix and expansion flexibility in production systems to understand and handle these challenges.

3.2.1. VOLUME FLEXIBILITY

The key ratio of the volume flexibility shall evaluate the possibility of a short-term, profitable capacity adjustment without a change of the amount of elements and the structure of the examined production-system. The evaluation of this kind of flexibility is conducted by looking at the marginal, quantitative output within a predefined period of time (e.g. a week or a month). In this period, a profitable production is feasible. Hereby the relevant key ratios applying to the defined observation period are the break-even-point and the capacity maximum of the production-system. While the break-even-point is dictating the output whose earnings exactly cover the relevant expenses, the capacity maximum is marking the greatest, still profitable (quantitative) output of the production system. The capacity maximum is determined by the technical efficiency and organisational steps (e.g. overtime or changes of the shift work). Both ratios fix the limits of a range called “Flexibility Space”. Within this Flexibility Space the production is economically flexible (cp. figure 3).

![Figure 3: Evaluation of the Flexibility Space in a production system](image)

3.2.2. MIX FLEXIBILITY

The concept of the Mix flexibility intends to assess the freedom in composite of the product/variant mix of production programme. The freedom is the ability to renounce certain products or to substitute them without an impact on the highest reachable profit. Starting with an optimal production-programme, it makes sense to determine the Mix flexibility by using the average deviation of the production-profit. This deviation is derived from the quantitative change of the products that need to be manufactured. The necessary ratios for that are the system-optimal production-profit and the product-specific profit-deviation.

The system-optimal production-profit is the greatest possible profit for a certain period of time that can be reached when the production-programme is composed optimally. This involved the type of products as well as the amount of products to be manufactured. On the contrary, the product-specific profit-deviation is investigating how a quantitative change of a single product influences the system-optimal production-profit while all other circumstances remain unchanged. Consequently the highest possible profit of the examined production-system needs to be calculated, where as a single product out of the manufacturing portfolio is not being manufactured. This calculation needs to be conducted for every single product of the manufacturing portfolio.
profit-deviation is derived from the difference to the system-optimal production-profit. In order to be comparable, heterogeneous production-systems are to be described in the form of a ratio. By computing the arithmetic mean for out of all product-specific profit-deviations contained in the production-system, it is possible to determine the average profit-deviation of the production. According to this concept a production-system is completely mix-flexible, if the profit always remains constant, no matter what product out of the manufacturing portfolio is being manufactured.

3.2.3. EXPANSION FLEXIBILITY

The expansion flexibility shall show the ability of a production-system to increase its capacities permanently by changing its amount of elements and/or its structure. Due to the fact that it describes a system-specific Cost-Benefit Ratio with regard to the capacity increase, the economical expenditure for a capacity increase is the crucial evaluation parameter. The difficulty hereby is the quantification of this ratio. In addition to that, often there are various alternatives of increasing capacities. Those alternatives exhibit different intensity of times and costs and can lead to differential increases of the capacities. Therefore, a so-called target-capacity must be fixed. The target-capacity states by how many percentages the current Capacity Maximum has to be increased. Through that the alternatives of expansions which do not reach the new border of capacity can already be eliminated. All other alternatives that each the target-capacity respectively surpass it, are quantified on the basis of a homogeneous evaluation method by using the quantitative Flexibility Space explained in section 3.2.1. Hereby the previously defined target-capacity is assumed to be a fix value: the new Capacity Maximum of the system. Notwithstanding whether there is one or more potential alternatives surpassing the target-capacity. So the size of the Flexibility Space of each permissible alternative of expansion depends on its break-even-point. The earlier an alternative reaches this point, the greater is its Flexibility Space and consequently has a more positive effect on the Cost-Benefit Ratio of the examined system. Finally, in order to determine the Cost-Benefit Ratio in a quantifiable way respectively the economic expenditure resulting from it, the best alternative of expansion (the one with the biggest Flexibility Space) is being compared with the original Flexibility Space of the system. The proportional deviation of the two Flexibility Spaces from each other gives information about the expansion flexibility of the system (cp. figure 4).

![Figure 4: Valuation of the Cost-Benefit Ratio of an expansion alternative](image)

3.3. BASIC CALCULATIONS USED IN ecoFLEX

Based on the conceptual considerations of the assessment methodology, the flexibility relevant parameters that determine the volume, mix and timing are dependent on cost and product-dependent restrictions. Thereby, because of the common existence of various manufacturing facilities of one or more products, for example a number of existing identical or similar resources an manufacturing flows, these parameters could ultimately lead to
inconsistent or even contradictory assessments of flexibility. Therefore, it is absolutely necessary, to determine the relevant parameters on the basis of their limits (minimum or maximum), based on a valid production plan and the current restrictions, that are thus optimal. Such a production plan, often called production program, describes the system resources in terms of their nature and extent of their use for a specified period, leading to a certain number of manufactured products.

The common challenge in the calculation of the various flexibilities within the valuation methodology is thus the determination of restriction conform, optimum production programs, to identify economic production limits. This leads to so-called optimization problems, which can be adopted as linear. Their solution is usually using the simplex algorithm, which should be used as the main method of linear optimization problems. Before the simplex algorithm can be applied, however, it is first necessary to standardize the optimization problem. This is achieved through the ecoFLEX-specific basis algorithm, which is described by the following three successive steps.

3.3.1. FORMULATION OF THE OBJECTIVE FUNCTION (STEP 1)

With the formulation of the objective function, result variables will be defined in the form of a vector $x \in \mathbb{R}^{|R|}$, which describes a production plan. It defines to each product EZG and workplace AP the manufactured production quantity $x_{EZG,AP}$. Thereby $R$ indicates the amount of all product-workplace combinations, $E$ the set of all products and $W$ the amount of all workplaces. An objective function is required for the calculation of the result variables, in order to resolve the respective optimization problem in terms of identifying an optimal production plan. It is represented by a vector $c \in \mathbb{R}^{|R|}$, where $R$ is the set of all possible product-workplaces combinations. The target value of this function $c$ for a determined production plan, is the scalar product of $c$ and $x$:

$$c(x) = c^T \cdot x = c_1 \cdot x_1 + c_2 \cdot x_2 + \ldots + c_{|R|} \cdot x_{|R|}$$

Each vector component $c_i$ represents the target value, which stands for the manufacture of the products EZG at workplace AP. By the calculation of the individual characteristics of amounts-, mix- or extension-flexibility this objective function may vary. This depends on the criteria which identifying the particular characteristic.

3.3.2. FORMULATION OF THE CONSTRAINTS (STEP 2)

The constraints, expressed by the mathematical and logical relations of the respective optimization problem, as well as the target function are depending on the flexibility relevant indicator, which has to be determined. Therefore they can vary. However, within the basic algorithm two types of secondary conditions can be provided. They are valid for all in the valuation methodology feasible and flexibility relevant calculations.

The time conditions are the first type of constraints. They represent the main limiting factors for the relevant optimization problems. They result from the fact that the production of one unit of the product EZG at workplace AP a certain process time $t_{PZ,EZG,AP}$ is needed. Moreover, the maximum working time for each workplace $t_{max,AP}$ per period $P$ is constrained by the used working time model $AZM$ and the aditious workplace-specific idle periods $t_{zli,AP}$. As a consequence an inequality constraint reveals for each production system contained in the workplace, which can be summarized as a matrix inequality:

$$\widetilde{T}x \leq T_{max}$$

$\widetilde{T}$ stands for a matrix of size $|W| \times |R|$. Each value $T_{i,j}$ of the matrix $\widetilde{T}$ represents the process time, which workplace $AP_i$ requires to produce the product from $r_{i,j}$. If $r_{i,j}$ doesn’t apply to the workplace $AP_i$, then the respective matrix entry assigns the value 0 as a process time. On the other hand, $T_{max}$ describes a vector, which indicates the maximum operating time $t_{max,AP}$ minus the extra idle costs $t_{zli,AP}$ for every workplace $AP$. 


The **ratio conditions** are the second fundamental type of constraints of the basic algorithm. They result from the product mix, which has to be specified for the flexibility calculations and from the component dependencies of the individual products. The latter can be described using a so-called component matrix \( C \in \mathbb{R}^{n \times n} \), which specifies how many units of a product are directly used to manufacture a unit of another product. The product mix, which represents the ratio of finished products in terms of their sale, is described with the vector \( v \in \mathbb{R}^{|P|} \). For not saleable products \( EZG_i \in ZP \) applies \( v_i = 0 \). The simplex algorithm provides the manufactured product amounts to be sold, which following this vector, have to be a multiple of \( v \). That means the saleable output \( x'_i \) and \( x'_j \) of two products \( (EZG_i, EZG_j) \in E \) must meet the restriction \( x'_i : x'_j = v_i : v_j \). Since according to the matrix \( C^{n \times n} \) self-manufactured products can be used in other products and because of the inclusion of defective products, it is quite possible that the total output quantities differ from this proportion.

In order to describe the ratio conditions resulting from the product mix in a simple mathematical form, a ratio matrix \( V \in \mathbb{R}^{|P| \times |P|} \) has to be defined, summarizing all product mix relations. In addition, it has to detach any units from the relation, which are no longer available because of defect or further processing. This requires a priori knowledge of the respective total production \( x_{EZG} \) for each product \( EZG_i \) on all workplaces \( AP \). On the other hand, an a priori knowledge of the overall needs for every product \( EZG_i \in ZP \) used for the manufacture of downstream products in production process is also necessary. The resulting calculations are summarized in the base algorithm represented in the following matrix equation:

\[
V \cdot x = 0
\]

This matrix equation leads to an optimal production plan. However, it doesn’t have to exploit all production capacity, as the product mix has to be taken into account. The limiting factor is thus the product, which reaches its capacity limits first.

### 3.3.3. FORMULATION OF THE LINEAR OPTIMIZATION PROBLEM (STEP 3)

In order to formulate the basic optimization problem for calculation of various flexibilities in ecoFLEX, it is necessary to convert the matrix equation, which indicates the ratio conditions. This necessity arises from the fact, that the standard procedure for solving linear optimization problems doesn’t allow linear equations as a constraint. A possibility is to transform each of the equations into two inequalities, as an equation \( a = b \) is equivalent to \( a \leq b \wedge (−a) \leq (−b) \). The resulting inequalities led during the development of ecoFLEX to numerical difficulties, so that often an invalid solution in the usual data types was displayed.

For this reason, a personal innovative extension of the linear optimization problem was made. It provides an efficient reformulation of the ratio conditions. The general idea is that each of the equations in the summarized ratio matrix makes it possible to calculate a solution variable for the other matrices. Thus, it is not absolutely necessary to determine all variables in the course of solving the optimization problem. It is better to use the simple weighted sum that results from the given equations. This will prevent the mentioned numerical problems and simplify the solution, because the algorithm must consider less variables than previously. The applied calculation rule is explained as follows:

The equation system given by the equation \( V \cdot x = 0 \) should be solved according to the appropriate variables. This results in a form in which these variables can be easily calculated by inserting the values of all other variables. The coverage of these variables using the vector \( y \), allows the formulation of a transformation matrix \( E \), so that \( x = E \cdot y \). All calculated quantity vectors fulfill the ancillary condition equation, and vice versa, all the quantity vectors, which fulfill the ancillary condition, can be represented in this form. As a consequence of this reformulation it is necessary to transform the objective function and the remaining constraints too, so that only a linear optimization problem \( y \) has to be solved. Summarized results thereof:
### Requirements

Search for production plan \( x = E \cdot y \), which lead to a maximum objective function.

### Objective Function

\[
e(y) = c^T \cdot Ey = (c^T E) \cdot y
\]

### Constraints

- \( \tilde{T} \cdot Ey = (\tilde{T} E) \cdot y \leq T_{\text{max}} \)
- \( x = Ey \geq 0 \)

![Figure 5: Valuation of the Cost-Benefit Ratio of an expansion alternative](image)

Using this clever transformation of the constraint conditions in the form of linear inequalities is the applied mathematical model for the basic algorithm, consisting of calculation parameters, outcome variables and constraints, transformed in a form from which with the help of standard solution procedures an optimal production plan can be determined.

### 4. SUMMARY

This article provided a detailed insight into the procedure of flexibility measurement of production systems using the ecoFLEX method. This approach allows a novel, comprehensive manner of examining and evaluating flexibility. Th ereby various levels of the system are considered, from the workstation to the factory level. Simultaneously, it creates a fundament for the comparability of production systems of different industry sectors. That facilitates companies to unerringly incorporate the flexibility as a decision criterion when selecting and forming the production-system suitably.

Currently the evaluation method ecoFLEX is experimentally applied by a reputable, globally operating automobile supplier. One of its production areas is analyzing deficits in flexibility and evaluating alternatives of extensions on the basis of actual production data. Furthermore, the evaluation-method is assimilating additional requirements regarding possible analyses. By doing so, it is pushing the expansion and itemisation of the functionality and of the applicability of ecoFLEX.

### REFERENCES


Beneficial Re-use of Waste Foundry Sand: argument for facilitated Industrial Symbiosis

Kirk Bridgewood1, Martin. G. McKie1
1School of Science and Technology
University of Teesside
Middlesbrough TS1 3BA, UK

ABSTRACT

Within the UK it is estimated that around 1 million tonnes of Waste Foundry Sand (WFS) is produced, of which approximately 300k tonnes goes to re-use applications. Such applications include road embankment construction, ground improvement, concrete manufacture, flowable fills and hydraulic barriers or liners. In this study, WFSs from an iron and steel foundry were analysed for physical and chemical properties including: particle size distribution; organic matter content (OM); pH; electrical conductivity (EC); total carbon (C) and nitrogen (N) content; C: N ratio; available/soluble nutrients (mineral N, P, K and Mg); and heavy metals content to assess their suitability for manufactured soil and asphalt under the standards BS3882 and BS EN 13043 respectively. Further leachate analysis and hazardous waste analysis was carried out in accordance with the Waste Acceptance Criteria. The results showed that despite the initial fear that the heavy metals, phenolics and Poly Aromatic Hydrocarbons (PAHs) present in WFS could limit its application to soils; the limiting factor was the grain size distribution of the sand which led to the finer fraction being used in asphalt production. The total analysis of the WFS alone demonstrated very low levels of heavy metals which were within the Soil Guideline Values and within BS3882. The analytical results were used to facilitate the re-use of the WFS for both manufactured soils and asphalt. The flows of materials are illustrated and the combination of reduced economic and environmental costs provided a strong argument for the development of symbiotic relationships.

1. INTRODUCTION

Numerous waste materials are generated from manufacturing processes although the increasing environmental awareness has contributed tremendously to concerns related to their disposal. With the scarcity of space for landfilling, and its ever increasing cost, waste utilisation has become an attractive alternative to disposal [1]. The issue is addressed increasingly by alternative scenarios of beneficially reusing Waste Foundry Sand (WFS). Beneficial reuses of WFS span a variety of applications related to infrastructure engineering and rehabilitation works, e.g. road embankment construction [2, 3], ground improvement [4], concrete manufacture [5], flowable fills [6], hydraulic barriers or liners [7]. The net results are cost savings for both foundries and user industries, and an environmental benefit at the local and national levels.

In the UK, around one million tonnes of WFS are produced annually [8], of which 70% is believed to be sent to landfill. Since the introduction of the Landfill Directive (1999/31/EC) [9], which has been in force since June 2002, the cost of landfill has risen considerably with current landfill tax set at £32 per tonne and set to increase by £8 annually until at least 2010. When gate fees and transport costs are taken into consideration this figure is close to £60 per tonne. Besides the financial burden to the foundries, landfilling WFS also makes them liable for future environmental costs, remediation problems and regulation restrictions [10].

This study was made out as part of the National Industrial Symbiosis Programme1 (NISP), and the methods and results present an argument for facilitated Industrial Symbiosis (IS). Research made at the University of Teesside, under NISP was made to investigate the physic-chemical properties of a WFS from a North East foundry producing around 10k tonnes per annum. The results of the analyses were used to facilitate the re-use of WFS between three companies, as a substitute for virgin sand in the manufacture of top soils for the construction industry; and as a component of asphalt production to replace virgin sand.

1 Corresponding author: Tel.: (0044) 1642-384670; E-mail: k.bridgewood@tees.ac.uk
1 National Industrial Symbiosis Programme in the North East of England is co-ordinated by the Clean Environment Management Centre (CLEMANCE) at the University of Teesside and is funded through DEFRA.
2. WASTE FOUNDRY SAND

Metal castings are often made from a mixture of clean, uniformly graded silica sand, a binding agent (either clay or a chemical), water, and additives such as coal dust, cereal, fuel oil or wood flour [2] to produce a reducing atmosphere during the casting process to minimise defects [11]. The most commonly used chemical binders used to bond core sands (e.g. phenolic urethane, furan, novolac and resole) are made from basic components such as phenol, formaldehyde and furfuryl alcohol [12]. This mixture is referred to as foundry sand. Excess foundry sand is generated because varying amounts of these additives must be reintroduced continually to the sand to maintain its desired properties [3]. Therefore WFS can be defined as sand that can no longer be used to make metal casting molds and cores, which is due to changes in grain shape and size [13]. In addition excess foundry sands are also generated as a by-product of metal casting. The types of residues generated and the recovery process also present particularities. The variable characteristics need to be determined to find the most suitable application for reuse of the WFS.

Current regulations require that WFSs be tested using the Toxicity Characteristic Leaching Procedure (TCLP) under the Waste Acceptance Criteria [9] to determine if they are hazardous wastes. According to the TCLP test, many foundry sands (especially molding sands from ferrous foundries) do not qualify as hazardous waste and, apparently, present little threat to groundwaters [14,15]. Although, WFSs can be used to produce manufactured soils and other agricultural products [16]; however, potential WFS constituents, such as heavy metals and resins used to bind core sands, may impact soil microorganisms [17]. Elevated concentrations of heavy metals also render the material unsuitable for construction uses such as asphalt since they are classified as non-hazardous rather than inert. The main concern relating to casting sand is the presence of phenol, although concentrations in cast sand tend to be low. Most foundry sands may contain some PAHs resulting from incomplete combustion of organic constituents. Binders are, generally, harmful, toxic (H5/H6) and or irritant / corrosive (H4/H8) [9].

The beneficial use of WFS in manufactured or synthetic soils has gained recent notice [16, 18]. In these studies, the soil manufacturing process, generally involved the blending of low-grade soil with sand and an organic additive. Fine textured soils (e.g. silty clay) may benefit the most from WFS applications, as they often drain very slowly and remain wet for long periods [18]. Such soils are typically of construction sites. The addition of WFS may improve the saturated hydraulic conductivity ($k_s$) [13]. Because water movement is very important in high-foot traffic soils, large amounts of sand are used commonly in synthetic soils destined for putting greens and athletic fields [19] while manufactured soils are particularly useful for the construction industry. The study by Dungan et al. [13] concluded that increases in $k_s$ were seen in soils blended with WFS. The one exception was WFS from Non-Bake Sand (NBS) containing sodium bentonite as sodium bentonite is a swelling clay and could impede water movement. A study by Dungan et al [13] identified, that there is a need for studies that characterise the physical properties of soil foundry sand blends because of the interest in using WFS in synthetic soils. Previous studies have suggested that grain sizes from different foundry processes vary significantly based on the metals cast [20]. The physical characteristics also allow WFS to be used in asphalt. According to the United States Environmental Protection Agency [21] WFS can effectively replace conventional materials in hot mix asphalt mixtures without compromising the quality. The physical characteristics of spent foundry sand are similar to those of fine silica sand. Nearly all types of WFSs fall in the particle range between 0.1 and 0.6 mm. The uniformity of these by-products suggests their utility in manufacturing. In addition, asphalt production may benefit from the sub-angular shape of WFS if used as a mix ingredient [21].

3. INDUSTRIAL SYMBIOSIS

Industrial ecology proposes to see industrial systems (i.e. a factory, and eco-region, or a national or global economy) as being indistinct from the environment, but rather a unique type of ecosystem, based on infrastructural capital rather than on natural capital. In order for human industrial systems to be sustainable, they need to be modelled after natural systems in which waste is all reusable. Industrial symbiosis is directly related to industrial ecology and is concerned with the flow of energy and materials through regional economies; collaboration opportunities offered by geographical proximity is important and allows the user to avoid the high costs and impacts of transportation [22]. The by-products from one industry should be able to serve as a resource for another, ideally adjacent, industry.

Incorporating construction projects into the industrial ecology of a region requires a shift from an open loop system, referred to as a Type 1 Ecosystem [23] that utilises virgin resources and then disposes them at the end of their cycle, to one that utilises secondary materials from other industrial sectors such as a Type 2 Ecosystem [23]. Utilisation of industrial by-products (IBPs) helps to minimise impacts from mining and processing of virgin materials.
materials and disposal of IBPs [24]. Additionally, as industry is generally located in urban regions, the by-products are closely located to areas with higher roadway infrastructure needs and transportation of building materials can be minimised [24].

The benefits of synergies can be substantial and the positive impacts typically extend beyond the boundaries of the companies, covering the three sustainability areas: economic; environmental; and social (Table 1).

Table 1: Examples of sustainability benefits of regional resource synergies [25].

<table>
<thead>
<tr>
<th>Economic Benefits</th>
<th>Environmental Benefits</th>
<th>Social Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced resource input costs</td>
<td>Reductions of waste</td>
<td>Job creation</td>
</tr>
<tr>
<td>Reduced waste management costs</td>
<td>Reductions in resource consumption</td>
<td>Improvements in media quality, e.g. water and air</td>
</tr>
<tr>
<td>Development of new products and their markets</td>
<td>Reductions in waste generation</td>
<td>Improving the qualitative attributes of existing jobs, e.g. morale</td>
</tr>
<tr>
<td>Increased productivity of resources: water, energy; waste management; and logistics infrastructures</td>
<td>Reductions in pollutant emissions</td>
<td>Cleaner and safer environment for the local communities</td>
</tr>
</tbody>
</table>

Despite the many benefits outlined by Harris [25] many companies are still sending material to landfill. The barriers to IS are summarised by Gibbs et al [26] as follows:

- First, there are technical barriers, including the possibility that local industries do not have the potential to “fit together”;
- Second, informational barriers may make it difficult to find new uses for waste products, relating to poor information regarding the potential market and potential supply;
- Third, economic barriers may inhibit the incentive to use waste streams as a resource if there is no reliable market for them;
- Fourth, regulatory barriers may prevent industries or industrial processes being linked together; and
- Finally, there may be motivational barriers wherein firms and public sector agencies and other relevant local organisations must be willing to co-operate and commit themselves to the process.

4. METHODOLOGY

The WFS was analysed for physical and chemical properties, including: particle size distribution, organic matter content (OM); pH; electrical conductivity (EC); total carbon (C) and nitrogen (N) content; C: N ratio; available/soluble nutrients (mineral N, P, K and Mg); and heavy metals content for the purpose of BS3882 and BS EN 13043, and further leachate analysis and hazardous waste analysis were made in accordance with the Waste Acceptance Criteria [9]. The sampling method used was in accordance with BS EN 12579:2000, Soil Improvers and Growing Media – Sampling.

The resin used by the foundry is a Phenol urethane (PU) (IsocureR 389/689 [silica sand coated with 1.2% w/w resin]) with 0.1% Fe₂O₃, together with the gas catalyst (isocure GBX20) (Ashland Specialty Chemical Company). The process used to make the PU (Isocure) cores is known as shell [27]. The WFS does not contain sodium bentonite, and therefore should not impede water movement [13].

Figure 2 outlines a framework for identifying a by-product and the steps taken to reach an acceptable re-use option. These steps are based on methodology from NISP in addition to key steps taken from the Beneficial Reuse of Ferrous By-Products – Draft Guidelines [28]. The shading indicates the level of facilitation offered by NISP to achieve a symbiotic relationship.

The carbons savings associated with this synergy were calculated using the NISP carbon methodology approved by DEFRA. The savings are grouped into three categories listed below with the sources of the calculations are given in references. Input Savings – which reflect the lower embedded energy in recycled materials compared with virgin raw materials. Establishing embedded energy involves a high degree of uncertainty and assumptions about what would have happened in the absence of the synergy [29]. Process savings – carbon outputs resulting from a change
in the way businesses do things – reducing energy use or transport. These savings are relatively straightforward to establish and verify [30]. **Disposal savings** – where material is recycled or reused rather than disposed to landfill. Landfill gases and hence carbon equivalent emissions are reduced. These savings are relatively straightforward to establish, although, their timings can be spread over many years [29, 31].

5. **RESULTS AND DISCUSSION**

The revised processing of WFS at the foundry allowed green and core sand to be combined to be screened for metal fractions >4mm. Around 90% of the metal fraction extracted from the waste stream was then re-used in the foundry’s furnaces leaving 9000 tonnes of WFS for re-use. The combined waste stream was then sieved to split the WFS into two fractions, > 0.3 mm and <0.3 mm which corresponded to 30% and 70% of total waste stream, respectively (Figure 1). The National Industrial Symbiosis Programme organised a meeting between the waste management company and the foundry and this resulted in suitable equipment being sourced to carry out this process.

As shown in Figure 1 and Tables 2 and 3 the foundry sand was within the limits of heavy metals, PAHs and phenolic compounds set by BS3882 and soil guideline values. The concentrations of P, K and N were low and, therefore, would require around 30% (w/w) of the final product to be compost to raise the fertility value to BS 3882.

The WFS was within the basic standards for BSEN 13043 – **aggregates for bituminous mixtures and surface treatments for roads, airfields and other trafficked areas**. Based on the chemical analysis, the WFS could be classified as inert (Table 2) and therefore, acceptable for asphalt. In addition, the grading of the WFS allowed the material to be used for fine aggregate as shown in Figure 1 and, due to content of fines of <0.063mm not exceeding 3%, no further testing was required. Standard BS EN 13043 stipulates that any fine aggregate with >3% passing 0.063mm requires a methylene blue value (MBv) in accordance with EN 933-9 to determine what is referred to as ‘harmful fines’ (e.g., swelling of clay such as bentonite clay which is often found in foundry sands) content.
Table 2: Leachable component analysis of WFS compared with limit values.

<table>
<thead>
<tr>
<th>Analytical parameter results expressed as mg kg(^{-1})</th>
<th>WFS</th>
<th>Limit values expressed as mg kg(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Inert</td>
</tr>
<tr>
<td>As</td>
<td>0.38</td>
<td>0.5</td>
</tr>
<tr>
<td>Ba</td>
<td>0.45</td>
<td>20</td>
</tr>
<tr>
<td>Cd</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Cr</td>
<td>0.01</td>
<td>0.5</td>
</tr>
<tr>
<td>Cu</td>
<td>0.01</td>
<td>2</td>
</tr>
<tr>
<td>Hg</td>
<td>0.00008</td>
<td>0.01</td>
</tr>
<tr>
<td>Mo</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Ni</td>
<td>0.12</td>
<td>0.4</td>
</tr>
<tr>
<td>Pb</td>
<td>0.18</td>
<td>0.5</td>
</tr>
<tr>
<td>Sb</td>
<td>&lt;0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Se</td>
<td>0.001</td>
<td>0.1</td>
</tr>
<tr>
<td>Zn</td>
<td>2.7</td>
<td>4</td>
</tr>
<tr>
<td>Chloride</td>
<td>78</td>
<td>800</td>
</tr>
<tr>
<td>Fluoride</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Sulphate</td>
<td>516</td>
<td>1000</td>
</tr>
<tr>
<td>Phenol Index</td>
<td>&lt;0.1</td>
<td>1</td>
</tr>
<tr>
<td>Total Dissolved Solids</td>
<td>1580</td>
<td>4000</td>
</tr>
<tr>
<td>Dissolved Organic Carbon</td>
<td>708</td>
<td>500</td>
</tr>
<tr>
<td>Total Organic Carbon</td>
<td>1.2% (w/w)</td>
<td>3%</td>
</tr>
<tr>
<td>Moisture</td>
<td>&lt;2.0% (w/w)</td>
<td>&gt;6</td>
</tr>
<tr>
<td>Loss on Ignition</td>
<td>1.1% (w/w)</td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>10.1</td>
<td>&gt;6</td>
</tr>
</tbody>
</table>

Table 3: Total analysis of WFS compared to BS3882 / Soil Guideline Value Standard Limits.

<table>
<thead>
<tr>
<th>Analytical parameter results expressed as mg kg(^{-1})</th>
<th>WFS</th>
<th>BS 3882 and Soil Guideline Value Standard Limits</th>
<th>Analytical parameter results expressed as mg kg(^{-1})</th>
<th>WFS</th>
<th>BS 3882 and Soil Guideline Value Standard Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As</td>
<td>2</td>
<td>20</td>
<td>Acenaphthene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Ba</td>
<td>35</td>
<td>N/A</td>
<td>Acenaphthylene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Cd</td>
<td>0.2</td>
<td>30</td>
<td>Anthracene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Cu</td>
<td>9</td>
<td>200</td>
<td>Benzo(a)anthracene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Fe</td>
<td>10000</td>
<td>N/A</td>
<td>Benzo(a)pyrene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Pb</td>
<td>6</td>
<td>750</td>
<td>Benzo(b)fluoranthen</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Hg</td>
<td>&lt;0.1</td>
<td>480</td>
<td>Benzo(k)fluoranthen</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Ni</td>
<td>8</td>
<td>110</td>
<td>Benzo(g,h,i)perylen</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Mo</td>
<td>6</td>
<td>N/A</td>
<td>Chrysene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Se</td>
<td>&lt;0.5</td>
<td>8000</td>
<td>Dibenzo(a,h)anthracene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Zn</td>
<td>28</td>
<td>300</td>
<td>Fluoranthene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Available Mg</td>
<td>130</td>
<td>&lt;600</td>
<td>Fluorene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Available P</td>
<td>4</td>
<td>&lt;100</td>
<td>Indeno(1,2,3-c,d)pyrene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Available K</td>
<td>170</td>
<td>&lt;900</td>
<td>Naphthalene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Conductivity uS cm(^{-1})</td>
<td>690</td>
<td>1500</td>
<td>Phenanthrene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Nitrogen % (w/w)</td>
<td>&lt;0.01</td>
<td>&gt;15</td>
<td>Pyrene</td>
<td>&lt;0.1</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Organic Matter</td>
<td>1.1</td>
<td>5-20</td>
<td>Total PAH</td>
<td>&lt;0.5</td>
<td>78100</td>
</tr>
<tr>
<td>pH</td>
<td>8.2</td>
<td>&lt;8.5</td>
<td>78100 mg kg(^{-1}) based on 5% Soil Organic Matter for industrial use.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Technical barriers were a significant issue for this synergy. However, the fact that there were standards available to compare could be viewed as a benefit. The motivational barrier is often significant when the company identified to receive a material is not completely convinced of the benefit of the secondary waste product based on the properties of the material and its suitability to replace virgin materials. In this case, the standard allows a direct comparison and a guarantee that the material is fit for purpose. The facilitation though NISP to carry out the testing and gather the data on the relative standards and thereby overcome the motivational barrier greatly strengthened the synergistic relationship being forged. Further to this, the report produced from the analysis (figure 1 and tables 2 and 3) was presented to the Environment Agency to convince them of the beneficial use of a waste and transform it into a product and remove the regulatory barrier (overcoming barrier number four as outlined by Gibbs [26]). Motivational barriers were also overcome by NISP organising and facilitating all meetings between the three companies and liaising with the Environment Agency. Figure 2 highlights the collaboration between NISP and the three companies in stages 1, 2, 7 and 10. Although the final stages of contract negotiation were carried out by the companies themselves.

Informational barriers, as Gibbs [26] outlines as ‘the difficulty in finding suitable nearby uses’, is a significant barrier to IS but were overcome by facilitated IS. In this example, NISP removed this barrier by implementing the stage independently of the foundry company. Based on previous literature, uses for WFS were identified by the NISP team. Near by industry capable of utilising the WFS were sourced, in this case through existing NISP members. Information stored on the NISP database about the capacities and types of material the member companies were interested in allowed NISP to match and further initiate a synergistic relationship. This is shown in Figure 2 by stages 3, 8 and 9 as heavily facilitated.

Economic barriers were overcome by finding solutions that proved costs effective, and it is a stage shown in Figure 2. The cost savings of this synergy are shown in Table 4. Implementing a symbiotic relationship between three (figure 3) companies through NISP helped to reduce the economic and environmental impacts of the WFS (table 5). However, it is important to note that but not for the legislation barrier of landfill tax, the economic case would have been very much weaker. Landfill tax and the imminent pressure of set annual increases, motivate re-use solutions which until now have not provided costs savings. Therefore, a barrier such as regulatory barrier is actually an aid in terms of IS activity, and one that enables NISP to operate in such a wide spectrum of industry. In addition to the savings listed in table 5 based on the work by Harris [25], it must be assumed that there are social benefits as shown in Table 1. These assumptions can be quantified by the reduction in transport. This is not evident if Figure 3 although it is reflected by the total CO$_2$ savings of 2450 tonnes achieved by not sending the material to landfill and

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>Facilitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Waste assessment to determine current waste stream quantities and qualities</td>
<td>None or little facilitation</td>
</tr>
<tr>
<td>2.</td>
<td>Determine the current cost of disposal</td>
<td>Carried out by both NISP and individual companies</td>
</tr>
<tr>
<td>3.</td>
<td>Identify wastes that have a potential to become by-products</td>
<td>Carried out by NISP</td>
</tr>
<tr>
<td>4.</td>
<td>Segregation of by-products from each other and waste streams</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Representative sampling of by-products</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Analysis of samples for contaminant concentrations</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Determine consistency of by-product contamination over time</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Match by-product with these specifications</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Assess potential human health and environmental impacts from the by-products</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Determine reliability and means of supply over time</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Perform a cost benefits analysis to demonstrate a return</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Enter into agreement with re-user</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Continue segregation and on-going quality assurance</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Series of steps taken to achieve an acceptable re-use option [28].
the two companies not using virgin materials. Although the distance covered was higher by 1035 miles per annum, the replacement of virgin materials i.e. the metal fraction returned to the foundry, and the replacement in soil and asphalt had a greater carbon footprint, and, therefore, saving [31].

Table 4: Annual savings achieved by the synergy.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Saving through IS activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂ savings (tonnes CO₂)</td>
<td>2,450</td>
</tr>
<tr>
<td>Landfill diversion (tonnes)</td>
<td>9,800</td>
</tr>
<tr>
<td>Virgin materials saved (tonnes)</td>
<td>9,000</td>
</tr>
<tr>
<td>Cost savings (£)</td>
<td>320,000</td>
</tr>
</tbody>
</table>

Figure 3: Flows of materials for re-processing.

6. CONCLUSIONS

Many issues were raised during this study around the stringent standards of BS EN13043 and BS 3882. The WFS tested was very close to the limit values of the two standards and to being classed as inert for BS EN 13043. Based on literature of other foundry sands, very few would meet these requirements. However, by testing WFSs as in this study, it could be assumed that a re-use application could be identified. To replicate synergies such as this, further research is needed in the development of a classification system that consists of four or five categories based on contaminant threshold concentrations, with appropriate reuse options for each category. The reuse options would be dependant on the threshold concentrations and the risk involved in the particular reuse.

This study has also strengthened the positive argument for facilitated IS offered through NISP. Key stages in the synergy were dependent on input from the NISP team which led to significant environmental and economic benefits; outcomes that possibly would not have been achieved without the level of facilitation offered by the NISP team from CLEMANCE at the University of Teesside.

7. REFERENCES


Thermal Effects and Pressure Distribution in Regenerative Braking

S. Sarip*, A. J. Day, P. Olley & H. S. Qi

School of Engineering, Design and Technology
University of Bradford
Bradford, West Yorkshire, BD7 1DP, U.K

ABSTRACT

A brake which is designed to provide a limited proportion of the total braking force at normal decelerations may be
able to brake the car at high decelerations but may suffer damage when used for emergency braking. A conventional braking system uses 100% of friction braking to retard a moving car while with regenerative braking the friction brakes will supply less than 100%. This paper starts to establish the working limitations of lightweight brakes for regenerative braking in extreme deceleration. Thermo-mechanical analysis can be used to predict the temperature profile on disc surfaces and analytical and experimental methods can be used to predict the thermal braking performance to calculate temperature rises under emergency braking conditions and brake cooling.

The aim is to establish an analytical and experimental procedure to design a lightweight disc rotor for general use at high duty levels.

1. INTRODUCTION

There are many hybrid cars in the market today. Customers tend to look for higher quality of passenger cars and lower fuel consumption; the world is facing an energy crisis and oil prices are rising. The hybrid powertrain is a solution to reduce dependence on oil, and the first hybrid car was invented in 1905. Hybrid vehicles use regenerative braking to recoup some of the energy dissipated during braking [1]; some of the braking power during deceleration can be absorbed and fed back to an energy storage device. However, even with regenerative braking, friction brakes are required to provide the necessary high duty braking performance, e.g. in the event of an emergency stop being required.

The size (and weight) of a car’s disc brake depends upon the performance required, specifically the braking torque, energy dissipation, and the power. It is possible to generate sufficient braking torque from a smaller brake, but the weight of power may over-reload the brake and cause physical damage. This paper describes initial work to establish the design envelope for a lightweight high power brake for use in transportation braking systems.

The performance of a disc brake in terms of torque, temperature distribution and wear is affected by the nature of the contact and pressure distribution at the interface between the friction pad and the disc rotor. These types of problems have been studied and have shown that the interface pressure between the pad and rotor is seldom uniform even in full apparent contact [2]. Published research used finite element analysis (FEA) and experimental work to study the effects of pad compression, load distributions and backplate rigidity [3]. The compressibility of the friction material is one of the major factors to define the pressure distribution at the friction interface, and can vary from 500 to 1000 MPa. A lower compression modulus i.e. a softer friction material gives a more uniform pressure distribution at the friction interface. In small-strain analysis the compression (or Bulk) modulus is directly proportional to Young’s modulus and is given by \( E/(3(1-2v)) \) where \( E \) is Young’s modulus and \( v \) is Poisson’s ratio. Usually most friction materials have a compression modulus value of around 1500 MPa. Thermal effects in brake interface pressure distributions have been studied using FEA in two ways: bulk temperature effects and macroscopic thermal effects [4]. Bulk temperature effects include thermal expansion, wear and coefficient of friction, and macroscopic thermal effects include hot spots, bending and rotor surface damage at the friction interface. Contact pressure analysis is a fundamental part of the prediction of frictional heat generation between a disc brake pad and rotor [5][6].

* Corresponding author: Tel.: (0044) 1274-234093; Fax: (0044) 1274-234124; E-mail: sbinsari@bradford.ac.uk
Three values of Young’s modulus, 210 MPa, 2000 MPa and 15000 MPa and one type of backplate of thickness 5 mm has been used here to investigate the effect of friction material on the interface pressure by modelling it using FEA. 233 solid elements of eight-noded brick and six-noded wedge types, with a total of 408 nodes form a three dimensional model complete with pad and backplate assembly. Using a Young’s modulus of 210 MPa the pressure distributions were found to be more uniform than for Young’s modulus values of 2000 MPa and 15000 MPa.

2. FRICTION MATERIAL COMPRESSIBILITY

2.1 DYNAMIC STRESS ANALYSIS

The pressure distribution on a brake pad was analyzed using a brake friction pair FEA model as shown in Figure 1. The centre of the disc was constrained in the axial direction (Z axis) so the disc could rotate about the axis. The model of the disc brake assembly with two brake pads with backplates is shown in Figure 1; this model comprised 3273 solid elements and a total of 5549 nodes. A ‘surface to surface’ contact scheme was defined using the FEA sliding contact algorithm.

![Figure 1: Annular disc complete with pads and backplates](image)

In order to restrain the pads when the disc is rotating, one of the backplate edges of each pad was constrained by a ‘hinge’ to allow rotation only about the z direction. The disc was set to rotate one revolution in one second with a friction coefficient of 0.4. The material properties of the model are shown in Table 1. Three values of friction material Young’s modulus are shown in Table 2.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Density, $\rho$ (kg/m$^3$)</th>
<th>Young’s Modulus, E</th>
<th>Poisson’s ratio, $\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel (Disc brake)</td>
<td>7050</td>
<td>116 GPa</td>
<td>0.27</td>
</tr>
<tr>
<td>Friction material (Pad)</td>
<td>2620</td>
<td>210 MPa</td>
<td>0.29</td>
</tr>
<tr>
<td>Backplate</td>
<td>7850</td>
<td>210 GPa</td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Friction material Young’s Modulus</th>
<th>Back plate thickness (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>210 MPa (Resin Bonded Composite)</td>
<td>5</td>
</tr>
<tr>
<td>2000 MPa (Aromatic polyamide fibre)</td>
<td>5</td>
</tr>
<tr>
<td>15000 MPa (Sintered metal friction)</td>
<td>5</td>
</tr>
</tbody>
</table>
2.2. RESULTS OF ANALYSIS

The pressure distribution for the three values of friction material Young’s modulus is shown in Figure 2. It can be seen that the area of friction material in contact with the disc depends upon the friction material compressibility. A suitable value of Young’s modulus for the friction material is 210 MPa which gives more uniform pressure between pad and rotor; this will help to reduce pad wear and encourage uniform wear, more uniform and lower peak brake temperatures and maintain a friction coefficient.

The pad friction material is usually worn more on the leading end of the pad compared to the trailing end. This is caused by the pressure distribution between the pad and the rotor which is higher at the leading end compared to the trailing end, created by a moment about the abutment and the friction drag force.

![Figure 2: Pressure distribution for different values of friction material Young’s modulus](image)

### Table

<table>
<thead>
<tr>
<th>E (MPa)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>Lower friction material Young’s modulus gives more contact area and smoother pressure distribution.</td>
</tr>
<tr>
<td>2000</td>
<td>Increasing the friction material Young’s modulus means decreasing the contact area between pad and disc.</td>
</tr>
<tr>
<td>15000</td>
<td>Higher friction material Young’s modulus gives less contact area and more concentrated pressure and temperature distribution.</td>
</tr>
</tbody>
</table>

3. THERMOMECHANICAL ANALYSIS

FEA, using the ABAQUS program [7] was used to predict temperature distributions in a typical disc brake complete with a pair of pads with backplates and two pistons. A full disc brake assembly FE model of a typical solid and ventilated disc brake was created in ABAQUS as shown in Figure 3 comprising 3239 elements and 4720 nodes with linear hexahedral elements type C3D8T (8-node thermally coupled brick, trilinear displacement and temperature).
3.1. FE MODEL

The contact surface frictional behaviour was simulated with a wheel rotational speed of 6.28 rad/s and 10.47 rad/s with an initial disc temperature of 20°C. Frictional heat was generated by pressing the pads against the disc with a uniform pressure of 4 MPa on the piston with the solid disc, but 4 MPa on the whole pad backplate area for the ventilated disc. Table 3 shows the material properties of the solid and ventilated discs used in the simulation.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Material</th>
<th>Density, $\rho$ (kg/m$^3$)</th>
<th>Young’s Modulus, $E$</th>
<th>Poison’s ratio, $\nu$</th>
<th>Conductivity, $k$ (W/m.K)</th>
<th>Specific Heat, $C_p$ (J/kg.K)</th>
<th>Thermal Expansion, ($K^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grey cast iron (Disc brake)</td>
<td>7000</td>
<td>116 GPa</td>
<td>0.27</td>
<td>53.3</td>
<td>103</td>
<td>$1.04 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>Resin BC (Pad)</td>
<td>2620</td>
<td>210 MPa</td>
<td>0.29</td>
<td>2.0</td>
<td>1100</td>
<td>$1.61 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>Backplate</td>
<td>7850</td>
<td>210 GPa</td>
<td>0.30</td>
<td>32.0</td>
<td>595</td>
<td>$1.17 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>Piston</td>
<td>7887</td>
<td>210 GPa</td>
<td>0.30</td>
<td>32.0</td>
<td>595</td>
<td>$1.17 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

The temperature distribution in a ventilated disc was also analysed using the same criteria as for the solid disc to investigate the temperature time history for selected nodes on the disc surface.

3.2. RESULTS OF ANALYSIS

Predicted temperature distributions on the solid disc rubbing path are shown in Figure 4. The temperatures were analyzed at 3 nodes along a radial line on the disc surface. On the solid disc a pressure of 4 MPa was applied to the piston to press the pad against the disc for about 2 seconds. The result was generated as the temperature distribution time history along the disc rubbing surface, and clearly indicates that disc friction surface temperature increased and decreased during one complete revolution (Figure 5). As the pads and disc are in contact only over the friction interface part of the disc surface, the heat generated in the friction interface causes nodal temperatures to increase rapidly. After the rapid temperature rise, a steady drop of temperature on the same node is clearly shown as it leaves contact with the pad.

The same phenomenon occurs for nodes in the disc below the surface. Frictionally generated heat flows into the disc creating a smaller temperature rise. Heat was then conducted to the whole disc and through into the ‘top hat’ section, the hub, and the wheel. Temperatures increase uniformly but slowly as heat is conducted from the frictional heat source. A similar analysis was completed for the ventilated disc (Figure 6) with an actuation force of 45450 N (4 MPa × $1.14\times 10^{-2}$ m$^2$) and a disc rotational speed 10.47 rad/s (100 rpm).
Temperature distributions were also analysed using the same criteria with the ventilated disc to investigate temperature vs. time for selected nodes on the disc surface as shown in Figure 7.
4 BRAKE COOLING

Temperature measurements were obtained using rubbing thermocouples and fine gauge wire probe thermocouple type K (Nickel-Chromium vs Nickel-Aluminium) and connected to a TC-08 thermocouple data logger with 8 input channels. Rubbing thermocouples were used to measure temperatures as they are accurate, cheap and easy to install. Temperatures were measured on the outer disc surface with the thermocouple mounted on the middle of the rubbing path. The temperature of the disc was measured after loads were applied at the brake pedal.

Temperature measurements were made every second using a TC-08 thermocouple data logger. Results were obtained from experiments for a ventilated disc and a non-ventilated (solid) disc. Each experiment was run at three different speeds, 450 rpm, 800 rpm and 1200 rpm corresponding to typical vehicle operation speeds. Each speed used different loads; 70 N, 50 N and 30 N respectively. The tests were performed in still air with the disc rotating. The results are shown for rotational speeds 450 rpm, 800 rpm and 1200 rpm for ventilated and non-ventilated discs in Figure 8 below. As already known, disc cooling is higher at high rotational speeds, and the ventilated disc has higher cooling efficiency than the solid disc.

Figure 7: Ventilated disc rotates in 10.47 rad/s (100 rpm)

(Actuating force = 4 MPa × 1.14×10^{-2} m^2 = 45450 N)

Figure 8: Cooling curve of disc brake
The cooling rate of a brake disc rotating at constant speed is linear as explained by Newcomb [8] who calculated a typical overall cooling rate for a car disc brake of $3.5 \times 10^{-3} \text{ s}^{-1}$ unit in 1965. The average heat transfer coefficient of a ventilated disc was found to be higher compared with a solid disc as shown in Figure 9 below. The heat transfer is affected by the area of disc exposed to the air; more surface area will increase the value of surface heat transfer coefficient. The gradient can be related to the heat transfer coefficient of the disc using a locally 1-D model giving the following heat transfer coefficients as shown in Figure 10.

Figure 9: Cooling rate of solid and ventilated disc

![Figure 9: Cooling rate of solid and ventilated disc](image)

Figure 10: Average heat transfer for ventilated and solid disc

![Figure 10: Average heat transfer for ventilated and solid disc](image)

5. DISCUSSION

Analysis of contact pressure distributions at the friction interface of a disc brake using FEA has been presented. The results show good correlation for understanding the effects of pad compressibility for a system where a load is applied to the pad and the disc rotates one revolution in one second. Decreasing the Young’s modulus of the friction material (and thus increasing the compressibility) encourages more uniform pressure distribution and a suitable friction material. Young’s modulus value is 210 MPa for the steel disc study. The result is very similar for both ventilated and solid brake discs.

Solid disc results show a similarity with ventilated discs in temperature distribution. It can be seen that heat in solid and ventilated disc was spread by conduction to the centre (hat) of the disc. The temperature at three different
places was compared with experimental results, and the results show a close similarity. Accurate temperature predictions of solid and ventilated discs assist in designing high performance disc brake rotors for commercial vehicles.

The cooling analyses have shown how the disc rotational speed affects the temperatures generated in the brake during the braking, and emphasizes the importance of convective disc cooling. The cooling rate results were used to calculate the heat transfer coefficient for three different speeds of rotation; it increased by approximately 100% for a solid disc between 450 rpm and 1200 rpm. For the ventilated disc the corresponding increase in heat transfer coefficient from 450 rpm to 1200 rpm was approximately 50%.

6. CONCLUSION

This paper presents an initial investigation of the thermal and contact/pressure parameters of a disc brake which indicates how the design of a disc brake is influenced by these parameters. It has been shown that friction material compressibility is important in pressure distribution analysis of brake: a very high Young’s modulus encourages an uneven pressure distribution. Thermo-mechanical FEA has been used to predict brake torque, energy dissipation and transient temperature distribution. The results give the temperature and contact pressure which are key factors in designing a lightweight disc brake rotor. The cooling of a disc brake is shown to be strongly enhanced by a ventilated design and increases steadily with rotation rate. Future research will include thermal and mechanically induced stresses in brake hub, wheel, tyres and suspension components.

REFERENCES

Stresses in bulk solids in cone hoppers: numerical solutions to the 3-dimensional stress distribution problem, using circular arc geometry

J.C. O’Neill*, A.J. Matchett, and A.P. Shaw

School of Science and Technology
University of Teesside
Middlesbrough, Tees Valley TS1 3BA, UK

ABSTRACT

In a previous paper, a 2-dimensional model of stress distribution within bulk solids, with circular arc principal stress orientation, in a wedge hopper was developed [1]. The model worked in an orthogonal, curvilinear coordinate system coincident with the principal stress trajectories: \((x-\psi_0)\) space. This paper presents a model with similar assumptions in 3-dimensional \((x-\varepsilon-\theta)\) space. Stress distributions for cone hoppers with rotational symmetry are now the subject of analysis. Rotational symmetry is assumed through angle \(\theta\). Three principal stresses are defined \((\sigma_R, \sigma_\varepsilon, \sigma_\theta)\). This is achieved via two static force balances on an incremental element, and assumption of a relationship between principal stresses. The numerical solution presented allows specification of arc stress along a given surface. As discussed in a previous paper [1], if a cohesive arch is specified, then stresses at the upper surface of the bulk solid are determined by the model. This calculated overpressure could be assumed to represent a horizontal material surface. Minimum flow outlet diameters from this model have been compared to available data.

1. INTRODUCTION

 Granular materials, or ‘bulk solids’, can be defined as any material composed of many individual solid particles, irrespective of particle size [2]. Granular materials are used in a wide range of industries, including the medical, food, construction, chemical and manufacturing industries [3,4].

 To allow processing of such materials, storage is required. Containers are often cylindrical, and can range in size from capacities measured in grams to thousands of tonnes [2]. At the base of the silo the container walls will converge to at least one small opening. This hopper section allows the flow of the granular material to be directed to the next stage of the process. In Figure 1, Schulze [5] describes some common problems encountered during flow of granular materials, including arching, funnel flow, rat-holing, flooding, segregation, eccentric flow and vibration.

 Many of the problems indicated above are caused by poor design of the silo and hopper set-up. Knowledge of stress distributions within these granular materials is not only concerned with ensuring flow of material from hoppers: such knowledge is also required for mechanical design of the hopper silo walls [6,7,8,9]. A lack of consideration of internal stresses

* Corresponding author: Tel.: (0044) 7715824815; E-mail: james0neil@aol.com

Figure 1: Possible problems during the operation of silos [5]
can result in catastrophic failure [10].

The stress analysis model proposed in this paper seeks to predict distributions 3-dimensional cone hoppers. The expansion from earlier methods detailed in this paper will give a better understanding of the problem, improved design algorithms, ensuring reliable shell design and material flow. In previous papers [1,11,12,13] new models making use of circular arc geometry were presented. The works produced were based on a model developed originally used by Enstad [14]. Enstad’s work calculated stresses in one direction only – in the vertical direction. Li [15] also made use of a model based on circular arc geometry. The models created for these papers initially provided force balance equations in two dimensions [11], and subsequently were expanded into three dimensions [12,13].

2. Model Geometry & Force Balance Equations

Model geometry has been defined in previous papers for two-dimensional hoppers [1,11] and three-dimensional hoppers/silos [12,13] with rotational symmetry. An important addition to two-dimensional models is the angle of rotational symmetry, whereby three-dimensional stress distributions can be observed. Matchett et al [1] and the current paper are developments from these prior works, and now take account of curvature normal to the circular arc after Lame-Maxwell [1,16].

The assumptions used in the new model are listed below. Incipient failure is assumed, therefore inertial terms are not included.

- Principal stresses act over successive sections circular arc sections of radius $R$ [11].
- The arc under consideration cuts the wall at vertical height $x$ from the vertex, and intersects the wall at distance $r$ from the axis of rotation [11]; Figure 2.
- The incremental arc has a thickness of $\delta_\alpha$, which varies across the span of the arc with $\alpha$ [11,12]; Figure 3.
- Positions within the vessel/hopper are located by height at which the arc cuts the vessel wall $x$, and arc angle $\alpha$ [12]; Figure 2.
- In three-dimensional space there are three principal stresses acting: radial stress $\sigma_R$, arc stress $\sigma_\alpha$ and azimuthal stress $\sigma_\theta$ [13]. Radial and arc stresses are orientated along circular paths of radius $R$. Azimuthal stresses act on the incremental element shown in Figure 3, and are orientated normal to the page.
- Rotational symmetry is assumed through azimuthal angle $\theta$, shown in Figure 2 [12].
Figure 2 [12] shows the principal stress arc geometry. The cone hopper has half angle to the vertical $\alpha_1$ and $\alpha_2$ respectively (for symmetrical systems $\alpha_1 = \alpha_2 = \alpha$); a circular arc cuts the right-hand side wedge at a distance $x$ above the apex with radius $R$. Point A has coordinates in $(X,Z)$ space of:

$$X = R \cos \epsilon$$  \hspace{1cm} (1)

$$Z = R \sin \epsilon$$  \hspace{1cm} (2)

If an incremental element is considered cutting the right-hand side with vertical height $\delta x$, and at an angle of $\epsilon$ to the vertical with incremental angle $\delta \epsilon$ - see Figure 3 [1].

A detail of the incremental element is shown in Figure 3 [1]. Using the circular arc geometry initially set out by Matchett [11], a force balance on an incremental element can be completed. A sketch of the rotated incremental element is shown in Figure 4 [12].

3. STRESS DISTRIBUTIONS

The $R$-$\epsilon$ coordinate system is not orthogonal-curvilinear, as shown in Figure 3. The line of constant $\epsilon$ between the two arcs is $FG$. This must be considered when force balances are constructed. The centre point of the upper arc does not coincide with the centre point of the lower arc - the arc centre moves from point $O$ to $O_1$.

Figure 3 [1] shows arc radii at angle $\epsilon$ for curves at $x$ and $x + \delta x$. Lines $O_1FE$ and $OMCD$ are parallel, with distance $CD$ equal to thickness $\delta v$. $M$ is the normal projection from point $O_1$ onto line $OMCD$.

Therefore:

$$R = a_1 x$$  \hspace{1cm} (3)

$$\frac{\partial w}{\partial x} = a_1 + a_2 \cos \epsilon$$  \hspace{1cm} (4)

where

$$a_1 = \frac{\tan \alpha}{\sin(\alpha + \beta)}$$ \hspace{1cm} and \hspace{1cm} $$a_2 = 1 - a_1 \cos(\alpha + \beta)$$

A benefit of the circular arch approach results from defining an incremental element that is co-incident with the directions of principal stresses. Calculation of shear stress is therefore not required in the analysis. In Figure 3, principal stress $\sigma_R$ acts on surfaces $CF$ and $DE$. While $CD$ and $EF$ are normal to line $CF$, they are not normal to line $DE$, due to precession of the arc centre from $O$ to $O_1$. The radius from $O_1$ normal to $DE$ is at angle $(\epsilon + \delta \epsilon)$ to the vertical. Therefore the surfaces on which $\sigma_\epsilon$ acts as a principal stress must be curved, as shown in Figure 3, and the incremental element upon which the force balance is based will be $CD,E,F$. $\sigma_\epsilon$ can be defined as a major principal stress acting upon the curved surface between $(R(x),\epsilon)$ and $(R(x + \delta x),(\epsilon + \delta \epsilon))$. From the system geometry:
\[ R_2 = \frac{\partial w}{\partial x} \frac{\partial \psi}{\partial x} \]  
(5)

and

\[ \frac{\partial \psi}{\partial x} = \frac{a_2 \cos \epsilon \sin \epsilon}{a_1 x \cos \epsilon} = \frac{a_2 \sin \epsilon}{R} \]  
(6)

Therefore

\[ R_2 = \frac{R(a_1 + a_2 \cos \epsilon)}{a_2 \sin \epsilon} \]  
(7)

For the cone hopper model, force balance equations are required to allow calculation of stresses, including those in the third dimension – azimuthal stress \( \sigma_\theta \). Three-dimensional force balances on the incremental element give equations in \( R \) - and \( \epsilon \) -directions.

\[
\frac{\partial F}{\partial \epsilon} = \left( \frac{\partial w}{\partial x} \sigma_\theta - F \right) \tan \epsilon - \sigma_R a_2 \sin \epsilon \right) + R \left( \frac{\partial w}{\partial x} \right) \rho \sin \epsilon
\]  
(8)

\[
\frac{\partial}{\partial x} \left[ R^2 \sigma_R \right] = RF - R^2 \left( \frac{\partial w}{\partial x} \right) \rho \cos \epsilon - 2R \sigma_R a_2 \cos \epsilon - Ra_2 \sin \epsilon \frac{\partial \sigma_R}{\partial \epsilon} + R \left( \frac{\partial w}{\partial x} \right) \sigma_\theta
\]  
(9)

Azimuthal stresses are found via use of the Haar-von Karmen hypothesis [2], or by other relationships. These relationships can be assumed to follow the form of \( \sigma_\theta = f(\sigma_\epsilon, \sigma_R) \). Using equation 9, it can be shown that when \( \epsilon \) is equal to zero, azimuthal stress is equal to arc stress. The relationship shown in equation 10 has been used for solutions demonstrated in this paper.

\[ \sigma_\theta = \sigma_\epsilon + k \epsilon \sigma_R \]  
(10)

### 3.1. Cohesive Arch Modelling

Radial stress \( \sigma_R \) values are set at zero at a position chosen to represent a cohesive arch location. This location can be provided by on-site data or by estimation using Nedderman’s equation 10.8.2 [2, p296]. Stresses at the top of the hopper are not fixed. Boundary conditions are obtained for arc stress \( \sigma_\epsilon \) by use of the Mohr-Coulomb criterion. This Mohr-Coulomb relationship is used only to provide initial values – stress distributions throughout the model are system are specified by the model. Azimuthal stress \( \sigma_\theta \) is obtained by a relationship with the other two principal stresses. After Matchett [12], azimuthal stress values can be calculated directly from equation 9. A boundary condition is needed, for example the relationship shown in equation 10.
Stress in bulk solids in cone hoppers: numerical solutions to the 3-dimensional stress distribution problem, using circular arc geometry

In Figure 5a, $\sigma_R$ values can be seen to increase from zero at the assumed location of cohesive arch. Radial stresses show a large overpressure at the material surface.

In the model used in this paper, it is assumed that a cohesive arch will be present when conformity to the Mohr-Coulomb criterion is demonstrated across the model. The initial height $H$ of the material surface has been reduced - Figure 5b shows that the results conform to Mohr-Coulomb limits, with the exception of a small number of results (shaded cells at base of figure). Increase of the $k$-value further increases stability. From the hopper geometry given in Table 1, the results of these figures equate to a critical diameter of 0.05 metres. A smaller hopper outlet diameter than this critical dimension will be subject to arching. Nedderman’s equation 10.8.2 gives the critical diameter as 1.22 metres.

3.2. ACTIVE STRESS STATE MODELLING

Radial stress $\sigma_R$ values are set to zero at a position representing the surface of the granular material. Again the Mohr-Coulomb criterion is used to produce an initial value for arc stress $\sigma_\alpha$. Azimuthal stress $\sigma_\theta$ values are specified by the same relationship given by equation 10. Model values are not restricted, other than active stress state relationship along hopper centre-line.

In Figure 6a, $\sigma_R$ values increase along the hopper centreline, towards the theoretical apex. This is in contravention to other models [14], where zero or negative stresses are assumed to indicate cohesive arching. It can be argued that if material within a hopper is stable, then stresses will increase to some positive value as per Janssen’s equation [2]. In Figure 6b partial conformity to the Mohr-Coulomb is demonstrated.
If a passive stress case is used, with non-zero radial stress values at the hopper top surface, then a decrease in radial stress values is demonstrated. Use of non-zero values at this position represents material above the hopper – for example during a typical hopper and silo arrangement. Stress distributions produced can be favourably compared to results presented by Enstad [14].

4. DISCUSSION

Stress distributions within granular materials provide useful information for hopper and silo design. Cohesive arch location can be predicted and avoided. Active and passive stress cases can be modelled, and therefore stress situations unviable to the Mohr-Coulomb criterion can be determined. Azimuthal stresses within hoppers can now be modelled to a level not previously possible.

However, there are limitations of \((x - \varepsilon - \theta)\) model. Boundary conditions may only be specified at one boundary in \(x\) and one in \(\varepsilon\). This means that if stresses are fixed at the bottom of the hopper (for example a value of zero \(\sigma_r\) representing a cohesive arch), then the surface overpressure is specified by the model. Two possible solutions are by assuming a material surface affects results, at the transition from open surface to circular arc principal stress orientation, or by introducing elastic effects throughout the system between boundary conditions at either end of the model. Alternatively, if it is assumed that the hopper will be placed underneath a silo, then results can be compared to previous models [2,14,17], which demonstrate peak stress values at the transition from silo to hopper.

There is a lack of data for comparison with model results. At the time of writing it is not possible to verify the relationship proposed between principal stresses, as no experimental data are available on stress distributions. Some data are available for critical outlet widths [2]. When compared with model data substantial differences in predicted outlet sizes were present. Jenike’s methods have been tested in industry; however some works [11] have indicated that an over-design may be present in the equation used. The geometry of the hopper should also be considered – a hopper of 1.2 metres in height and 2.1 metres in width, with a 1.22 metre outlet is unlikely to be susceptible to cohesive arching.

5. CONCLUSIONS

A three-dimensional model of stress distributions within cone hoppers has been presented, making use of rotational symmetry. The model provides radial, arc and azimuthal stress solutions through circular arc principal stress orientation. The information produced by the model can be used both for prediction of cohesive arch location and structural design of hoppers and silos.

The stress distributions produced have been compared to limited data. Model development would benefit from comparison to experimental data for verification of findings. Further work will include inserts and non-symmetrical hopper shapes.

Mathematical study of stress distributions within hoppers and silos is not a new discipline, however processing of granular materials in this way remains problematic [18].

REFERENCES


Stress in bulk solids in cone hoppers: numerical solutions to the 3-dimensional stress distribution problem, using circular arc geometry


**NOTATION**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>arc radius constant</td>
<td>[-]</td>
</tr>
<tr>
<td>a₂</td>
<td>arc thickness constant</td>
<td>[-]</td>
</tr>
<tr>
<td>g</td>
<td>acceleration due to gravity</td>
<td>[m/s²]</td>
</tr>
<tr>
<td>F</td>
<td>model variable, ( F = \sigma_e \left( \frac{\partial w}{\partial x} \right) )</td>
<td>[-]</td>
</tr>
<tr>
<td>H</td>
<td>value of ( x ) at for boundary condition in ( \sigma_R )</td>
<td>[m]</td>
</tr>
<tr>
<td>J</td>
<td>material ratio of effective stresses ( J = \frac{\sigma_e + T}{\sigma_{Re} + T} )</td>
<td>[-]</td>
</tr>
<tr>
<td>MCYF</td>
<td>Mohr-Coulomb Yield Factor</td>
<td>[-]</td>
</tr>
<tr>
<td>( r )</td>
<td>distance OA</td>
<td>[m]</td>
</tr>
<tr>
<td>( \bar{r} )</td>
<td>radius of rotation of incremental element</td>
<td>[m]</td>
</tr>
<tr>
<td>R</td>
<td>principal stress arc radius</td>
<td>[m]</td>
</tr>
<tr>
<td>R₁₂</td>
<td>upper arc radius</td>
<td>[m]</td>
</tr>
<tr>
<td>T</td>
<td>material tensile parameter – linearized yield locus</td>
<td>[Pa]</td>
</tr>
<tr>
<td>x</td>
<td>height of intersection of arc with Wall 2</td>
<td>[m]</td>
</tr>
<tr>
<td>X</td>
<td>vertical co-ordinate</td>
<td>[m]</td>
</tr>
<tr>
<td>Z</td>
<td>horizontal co-ordinate</td>
<td>[m]</td>
</tr>
<tr>
<td>( \alpha_1, \alpha_2 )</td>
<td>angle of wall to horizontal</td>
<td>[rad]</td>
</tr>
<tr>
<td>( \beta_1, \beta_2 )</td>
<td>angle of arc to wall normal</td>
<td>[rad]</td>
</tr>
</tbody>
</table>
\( \delta w \) incremental element thickness [m]
\( \delta \theta \) incremental change of arc centre O [m]
\( \delta \varepsilon \) increase in angle of orientation of stress – Figure 4 [rad]
\( \delta \psi \) angle between \( 0_1G \) and \( 0_1E \), due to progression of arc centres [rad]
\( \delta x \) incremental vertical height [m]
\( \varepsilon \) angular co-ordinate, angle between arc radius and vertical [rad]
\( \theta \) azimuthal angle [rad]
\( \rho \) bulk density [kg/m\(^3\)]
\( \sigma e \) arc stress [Pa]
\( \sigma_r \) radial stress [Pa]
\( \sigma_{\theta} \) azimuthal stress [Pa]

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>alpha</th>
<th>beta</th>
<th>Initial height</th>
<th>rho</th>
<th>phi</th>
<th>T</th>
<th>J</th>
<th>delta x</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>cohesive arch</td>
<td>30</td>
<td>5,5</td>
<td>1,2</td>
<td>1000</td>
<td>30</td>
<td>3000</td>
<td>3</td>
<td>0,006</td>
<td>3</td>
</tr>
<tr>
<td>active stress</td>
<td>30</td>
<td>17</td>
<td>2</td>
<td>1000</td>
<td>30</td>
<td>3000</td>
<td>0,33</td>
<td>0,005</td>
<td>1</td>
</tr>
</tbody>
</table>
A Bayesian approach for predictive maintenance policy with imperfect monitoring

Giuseppe Curcurù*, Giacomo Galante

Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale
Università degli Studi di Palermo

ABSTRACT

In the traditional preventive maintenance policy, the periodic maintenance activities are scheduled on the basis of the a-priori information about the failure behaviour of the population which the component belongs to, by assuming a probability distribution function and by estimating the involved statistical parameters. On the contrary, with the predictive approach, the maintenance activity is scheduled on the basis of the real degradation level of the component. So, it is possible to reduce the failure probability and, at the same time, to use the component for almost all its useful life. For this reason, the predictive maintenance policy makes possible the reduction of the maintenance costs with respect to the preventive approach and it is particularly effective for those components that must work with a high required degree of reliability in systems where failures can produce dramatic consequences. To apply the predictive approach, it is necessary to monitor the component degradation behaviour by using sensors. Nevertheless, before implementing a predictive policy, it is necessary to take into account the costs and the uncertainty of the monitoring system. In this paper we compare by simulation the effectiveness of the predictive maintenance policy with the traditional preventive one when the component must operate with a fixed reliability level. It is shown how the convenience of the predictive maintenance approach depends both on the parameters characterizing the stochastic degradation process and on the uncertainty of the monitoring system. For the preventive policy the a-priori information on the population is considered while, for the predictive one, this information is updated by a Bayesian approach using the data coming from the monitoring system.

1. INTRODUCTION

For some systems the main objective of the maintenance activities is the minimization of the failure probability or, at least, the preservation of this probability under a fixed value. In such systems the employ of sensors for the monitoring of their degradation level is very useful. In fact, this gives the possibility to follow the time history of the component and to identify the most appropriate time for maintenance activities, making possible the exploitation of the component for almost all its useful life. With the development of electronic monitoring systems and of modern diagnostic tools, it is increasing a significant interest for the predictive maintenance policy that, in some cases, is preferred to the traditional preventive one. In fact, the traditional preventive maintenance policy makes use of the a priori information on the population by assuming a probability distribution function and by estimating the involved statistical parameters. By a monitoring system it is possible to have further information on the stochastic degradation process of the particular component belonging to the population. The acquired information is then used to predict the remaining time to failure and to program the optimal time for the maintenance. This is important for people managing the maintenance area. It is, in fact, necessary to plan the maintenance resources and to prepare the plant for the maintenance activities with an adequate advance. Nevertheless, such sensors add new costs and exhibit inaccuracy in following the stochastic process. This inaccuracy implies an uncertainty in the supplied information. This occurs whether the degradation is defined as a geometric characteristic of the component or as the exhibition of a particular effect.

*Corresponding author: Tel.: (+39) 091-23861842; Fax: (+39) 091-7099973; E-mail: g.curcuru@unipa.it
For example, in a cutting tool, wear changes the geometrical characteristics causing an increase of superficial roughness on the machined parts [9].

If a maximum value of roughness is accepted, the condition of failed cutting tool corresponds to the reaching of such value. In this case, the vibration signal is not a correct fault indicator because it cannot follow the degradation process with correctness. A second kind of uncertainty depends by the monitoring system itself. In a monitoring system in fact, there are different parts: the link between the component and the sensor, the sensor itself, the transmission system, the signal amplifier, the acquisition system, etc.

For these reasons, a predictive maintenance policy presupposes the identification of a signal well correlated to the degradation process and a high precision monitoring system. In this paper, components whose sudden failure can produce dramatic consequences are considered. For this reason, they must operate with a high required degree of reliability [10]. It is assumed that the maintenance policy must assure a reliability level not lower than a fixed value. In order to choose between the implementation of the traditional preventive maintenance policy and the predictive one, a comparison is made by simulation. By hypothesizing a degradation model [1][8], it is shown how the convenience of the predictive maintenance approach depends on both the parameters characterizing the stochastic degradation process and on the uncertainty of the monitoring system.

To our knowledge, no paper deals with the uncertainty of the monitoring systems that we consider basically important for the choice of the maintenance policy. The paper is organized as follows: section 2 introduces the degradation model; section 3 the process monitoring; in section 4 the Bayesian approach for the updating of the a priori information is presented; section 5 shows the forecasting model and section 6 the proposed maintenance policy. Finally, section 7 reports the results of the simulation and some correlated comments.

2. DEGRADATION PROCESS

To apply the predictive maintenance policy, it is necessary to have reliable observations on the parameter of interest and to hypothesize a statistical model for the degradation process under study. We hypothesize, as degradation model, a first order autoregressive model with drift [9]:

\[ y(t + dt) = y(t) + \gamma' dt + \varepsilon(t) \]  \hspace{1cm} (1)

where \( y \) is the parameter that characterizes the wear, \( \gamma' \) is the mean value of the increment and \( dt(t) \) is a random error normally distributed with mean 0 and variance \( \sigma^2_\varepsilon \). In the model, the wear increments are assumed to be independent from the achieved wear level. If such dependence occurs, a multiplicative degradation model must be considered. Nevertheless, employing the logged value of the wear, it is easy to return to the proposed linear model.

Generally the degradation process is observed at regular time intervals \( \Delta t \). For this reason the discrete form of the model (1) will be considered. In this case \( \gamma = \int \gamma' dt = \gamma' \Delta t \). Considering \( \Delta t \) as the unit of time, the model becomes:

\[ y_{i+1} = y_i + \gamma + \varepsilon_{i+1} \]  \hspace{1cm} (2)

3. PROCESS MONITORING

The degradation path \( y \) can not generally be observed directly. The monitoring system supplies a parameter \( m \) correlated to the real degradation parameter \( y \). Let \( m \) be the value of such parameter at time \( t \). Supposed that the monitoring system has a linear response, the relation between \( m \) and \( y \) can be expressed by:

\[ m_i = a + by_i + \delta_i \]  \hspace{1cm} (3)

where \( a \) e \( b \) are the coefficients of the linear transformation and \( \delta_i \) is the total system error. Let \( \delta \) be normal distributed with mean 0 and variance \( \sigma^2_\delta \). From (3) \( y_i \) can be expressed as follows:

\[ y_i = \frac{m_i - a + \delta_i}{b} \]  \hspace{1cm} (4)

By substituting it in the equation (2):
A Bayesian approach for predictive maintenance policy with imperfect monitoring

\[
\frac{m_{i+1} - a - \delta_{i+1}}{b} = \frac{m_i - a - \delta_i}{b} + \gamma + \varepsilon_{i+1}
\]

(5)

Hence,

\[
m_{i+1} = m_i + b\gamma - \delta_i + \delta_{i+1} + b\varepsilon_{i+1}
\]

(6)

Equation (6) makes it evident that two values of \( m \) at two successive time instants are linearly linked. Setting the total error term as follows:

\[
\lambda_{i+1} = \delta_{i+1} - \delta_i + b\varepsilon_{i+1}
\]

(7)

equation (6) becomes:

\[
m_{i+1} = m_i + b\gamma + \lambda_{i+1}
\]

(8)

where \( \lambda \) is normally distributed with mean 0 and variance \( \sigma^2 = 2\sigma^2 + b^2\sigma^2 \).

4. HYPOTHESIS OF DIFFERENT DEGRADATION PROCESSES AMONG COMPONENTS BELONGING TO THE SAME POPULATION

In the model described above, \( \gamma \) is the mean drift that characterizes the degradation pattern of the components. It was considered as a known deterministic parameter, while \( \varepsilon \) represents the stochastic behavior of the degradation process. Therefore, the parameter \( \gamma \) represents a constant physical phenomenon common to all the units of the population. However, a specific component belonging to a population can exhibit a behavior that depends not only on the stochastic nature of the degradation process, but also on particular aspects, i.e. geometric or metallurgical characteristics or on different environmental working conditions. For these reasons, parameter \( \gamma \) must be considered as a stochastic variable and its value for a specific component as an outcome. In this paper a normal probability distribution with mean 0 and variance \( \sigma^2 = \sigma^2 \) has been assumed for parameter \( \gamma \).

The equations introduced in the previous section can be used on condition that each one is referred to a generic component \( j \) belonging to the population. Model (2) becomes:

\[
y_{j,j+1} = y_{j,j} + \gamma_j + \varepsilon_{j,j+1}
\]

(9)

The distribution of the stochastic variable \( m_{i+1} \), conditioned by the acquisition of \( m_i \), is normal with mean and variance given by:

\[
E[m_{i+1}] = m_i + b\mu_j
\]

(10)

\[
Var[m_{i+1}] = b^2\sigma^2 + \sigma^2
\]

Using data coming from the monitoring system, it is possible to estimate more accurately the degradation parameter \( \gamma_j \) for the \( j \)th component. As already said, such increase can be considered an outcome of the stochastic variable \( \gamma \) that expresses the different behavior of the components belonging to the same population. The initial available information on this outcome is the distribution of \( \gamma \), \( \pi(\gamma) \) with mean \( \mu \) and variance \( \sigma^2 \). By a Bayesian approach, the a-priori information \( \pi(\gamma) = \pi(\gamma) \) will be updated on the basis of the acquired data. Let \( m_i \) be the last acquisition of
the monitoring system, the a-posteriori distribution of $\gamma_j$ for the component $j$, $p(\gamma_j | m_i)$, is normal with mean $\mu_{\gamma,j}$ and variance $\sigma_{\gamma,j}^2$.

Setting, for sake of simplicity, the coefficients of the linear transformation $a=0$ and $b=1$, $\mu_{\gamma,j}$ and $\sigma_{\gamma,j}^2$ are given by [2]:

$$\mu_{\gamma,j} = \frac{\sigma_j^2 m_i + \sigma_j^2 \mu_j}{\lambda_j^2 \sigma_j^2 + \lambda_j^2}$$

$$\sigma_{\gamma,j}^2 = \frac{\sigma_j^2 \sigma_j^2}{\lambda_j^2 \sigma_j^2 + \lambda_j^2}$$

5. FORECASTING MODEL

The estimate of $m$ at time $i+1$, i.e. one time interval after the last acquisition, for component $j$ can be drawn from equation (8):

$$\hat{m}_{j,i+1} = m_{j,i} + b\gamma_j + \lambda_{j,i+1}$$

Hence:

$$\hat{m}_{j,i+2} = m_{j,i+1} + b\gamma_j + \lambda_{j,i+2} = m_{j,i} + 2b\gamma_j + \lambda_{j,i+1} + \lambda_{j,i+2}$$

By iterating the procedure, at time $i+k$,

$$\hat{m}_{j,i+k} = m_{j,i} + kb\gamma_j + \lambda_{j,i+1} + \lambda_{j,i+2} + ... + \lambda_{j,i+k}$$

The estimates of mean and variance of $\hat{m}_{i+k}$, given $m_i$, are:

$$\hat{\mu}_{m_{j,i+k}} | m_{j,i} = m_{j,i} + bk\mu_j$$

$$\hat{\sigma}_{m_{j,i+k}}^2 | m_{j,i} = b^2 k^2 \sigma_{\gamma,j}^2 + k \sigma_j^2$$

6. MAINTENANCE MODEL

The described monitoring system can be used for the implementation of a predictive maintenance policy. The maintenance policy must assure a reliability for the monitored component not lower than a fixed value.

The failure time distribution $f(\tau)$ of the population which the component belongs to can be used to compute its residual life. Let $F_{\tau}(\tau')$ indicate the unreliability of such component at time $\tau'$, $F_{\tau}(\tau') = P\{ \tau \leq \tau' \}$, i.e. the probability that the component failure takes place before $\tau'$. The monitored component will be considered failed when the degradation level will reach a limit value $y^*$ that corresponds to a threshold value $m^*$ for the monitored parameter.
Thus, the probability that the component fails before \( t^* \) is the probability that the degradation signal \( m(t^*) \) at time \( t^* \) is equal or greater than the threshold \( m^* \). Consequently, \( F(t^*) = P\{ \tau \leq t^* \} = P\{ m(t^*) \geq m^* \} \). Reliability \( R(t^*) \) is obviously \( 1 - F(t^*) \).

By using the forecasting model described in the previous section, mean and variance of \( m \) are updated at every signal acquisition by (15) and (16). It is so possible to estimate the distribution of \( \hat{m} \) at a generic time instant. If \( T \) indicates the necessary time to organize the maintenance activities, it is possible to estimate, at acquisition time \( t \), the degradation level of the monitored component at time \( t + T \). Assuming \( T \) as an integer multiple of the acquisition interval \( \Delta t \), the reliability of the generic component \( j \) at time \( t + T \), can be computed as follows:

\[
R_j(t + T) = \int_{-\infty}^{\infty} f(\hat{m}_{t+T}) d\hat{m} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{b^2 T^2 \sigma_{t,j}^2 + T \sigma^2}} \left[ \frac{1}{2} b^2 T^2 \sigma_{t,j}^2 + T \sigma^2 \right]^{\frac{1}{2}} \left[ \hat{m}_{t+T} - (m_i + bT \tilde{\mu}_{t,j}) \right]^2 d\hat{m}_{t+T} (17)
\]

In Figure 1, reliability at time \( t + T \) computed by (17), is represented by the area of the curve under \( m^* \).

7. COMPARISON BETWEEN PREVENTIVE AND PREDICTIVE MAINTENANCE POLICIES

To show how the parameters characterizing the degradation process and the monitoring system influence the advantages achievable with the implementation of the predictive maintenance policy, the degradation process of 3,000 components, following the stochastic process expressed by (9), was simulated using Matlab. The estimate of mean \( \mu \) and variance \( \sigma^2 \) of the population, that constitutes the a priori information, is obtained by the observation of the degradation behaviour of 25 components belonging to the population.

The a-priori information is used for the determination of the parameters of the failure time probability distribution and then for the evaluation of the maintenance interval, when a preventive maintenance policy is employed. The same information is updated by Bayesian approach when a predictive maintenance policy is adopted.

Considering the assumptions on the stochastic process, the distribution function of the failure time, under the condition that the threshold \( m^* \) has been reached, can be obtained by the Bayes formula as an a-posteriori probability. This distribution is known as inverse Gaussian [9]. From this distribution, it is possible to calculate the replacement interval \( t_{\text{prev}} \), respecting the constraint on reliability or equivalently the accepted risk. The same constraint on reliability is considered for the calculation of the replacement instant \( t_{\text{pred}} \) when the predictive maintenance policy is adopted. In both cases a failure risk of 0.01 is considered. In the case of predictive policy, \( T \) is fixed to 10, i.e. it is assumed that 10 time units are sufficient to organize the maintenance activities. It is also
assumed that the scheduled time for the starting of the maintenance activities can be changed only if a failure occurs in the time interval \([t, t+T]\).

For each simulated component, the times \(t_{\text{pred}}\) and \(t_{\text{prev}}\) are compared with the actual time \((t_{\text{real}})\) corresponding to the reaching of the fixed threshold \(m^*\).

The effectiveness of a maintenance policy is here evaluated taking into account its ability in determining the replacement time close to time \(t_{\text{real}}\).

As a measure of this effectiveness, the relative mean distances computed with the formulas (18) and (19) are considered.

\[
d_{\text{pred}} = \frac{1}{N} \sum_{j=1}^{N} |t_{\text{pred},j} - t_{\text{real},j}| / t_{\text{real},j} \quad (18)
\]

\[
d_{\text{prev}} = \frac{1}{N} \sum_{j=1}^{N} |t_{\text{prev},j} - t_{\text{real},j}| / t_{\text{real},j} \quad (19)
\]

The smaller the relative mean distance the better the maintenance policy considered.

Table 1 shows, for different combinations of the values of the variances \(\sigma_\gamma^2\), \(\sigma_\varepsilon^2\), \(\sigma_\delta^2\), the relative mean distances.

<table>
<thead>
<tr>
<th>Simulations</th>
<th>(\sigma_\gamma^2)</th>
<th>(\sigma_\varepsilon^2)</th>
<th>(\sigma_\delta^2)</th>
<th>(d_{\text{pred}})</th>
<th>(d_{\text{prev}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1835</td>
<td>0.2787</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1804</td>
<td>0.2749</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.1896</td>
<td>0.2787</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1813</td>
<td>0.2272</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.1945</td>
<td>0.1925</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.1891</td>
<td>0.1925</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>0.4</td>
<td>0.8</td>
<td>0.1995</td>
<td>0.1925</td>
</tr>
<tr>
<td>8</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1834</td>
<td>0.1925</td>
</tr>
</tbody>
</table>

Table 1: results of the simulations for the mean relative distances between times

These results show the influence of different sources of uncertainty on the adopted maintenance policy. Even if no experimental plan was projected, it is possible to draw some interesting considerations. The effectiveness of the predictive maintenance policy is little influenced by the variability of the parameter \(\gamma\), while it is more influenced by the variability of the degradation model \(\sigma_\gamma^2\) and by the uncertainty of the monitoring system \(\sigma_\delta^2\). On the contrary, the effectiveness of the preventive maintenance policy is more influenced by the variability of the parameter \(\gamma\) and little by \(\sigma_\gamma^2\). Obviously it is not influenced by \(\sigma_\delta^2\) because in the preventive maintenance policy no monitoring system is employed. It is so possible to conclude that the predictive maintenance policy is always better than the preventive one, excluding the cases with the highest values of \(\sigma_\delta^2\). The advantages of the predictive policy increases significantly with the increasing of the parameter \(\sigma_\gamma^2\), is less influenced by the variability of the degradation process and decreases with the increasing of the uncertainty of the monitoring system, until this policy becomes not convenient.
8. CONCLUSIONS

In this paper, a comparison between the traditional preventive maintenance policy and the predictive one is conducted by simulation. Different sources of variability are considered and their impact on the maintenance policy investigated.

It is shown that the predictive maintenance policy makes possible a better exploiting of a component taking into account its actual degradation level. With the increasing of the uncertainty of the monitored system, this advantage is gradually lost. The proposed procedure makes possible an early estimation of the convenience in the implementation of a predictive maintenance policy, in situations where the reliability of the system under study is the most critical parameter. The proposed Bayesian approach of integration between information coming from a monitoring system with the a-priori knowledge of the degradation process, when a predictive maintenance policy is adopted, will be the subject of a future work with the objective of minimizing the total maintenance costs.

References

Robust Maintenance Scheduling for Reliability Maximization

A. Certa*, G. Galante, T. Lupo and G. Passannanti

Dipartimento di Tecnologia Meccanica, Produzione ed Ingegneria Gestionale
Università di Palermo
Palermo, 90128, Italy

Abstract

The present paper aims to single out the maintenance actions to perform on a system constrained to be maintained only during some planned stops. Since system failure implies costs and risks for people and/or the environment, then it is necessary to minimize the probability of its occurrence. Thus, maintenance actions need to maximize the system reliability up to the next planned stop, in respect to some constraint. The originality of the problem discussed in the present paper lies in considering reliability values affected with uncertainty within a range of vagueness. Consequently, a further problem is added to the constrained reliability maximization, that is the search for a robust solution: the selected solution also needs to guarantee a low sensitivity to the real position of the component reliability within the variability range. This implies the formulation of some parameter able to express the solution robustness at best.

To solve the problem, the authors have developed an exact dynamic programming algorithm that is also suitable for complex series-parallel systems, usually intractable by the mathematical programming. The algorithm requires short computational times and thus quickly allows to individuate the optimal solutions for different sets of components reliability values, so constituting a valid decision support tool. Finally a numerical example applied to a complex system composed by a large number of components is reported.

1. Introduction

In the last few years, many researchers have faced some maintenance problems in the multi-component systems field. From a reliability point of view, these systems are constituted by several components in a series disposition, some of which are in their turn parallel or parallel-series components. A wide overview about the multi-unit system maintenance models developed up to 1991 is presented by Cho and Parlar [1], while a complete updated survey can be found in Dekker et al. [2] and Nicolai and Dekker [3].

In literature, the most used measure in evaluating the maintenance policy performance is the stationary availability of a production system. The various models differ for the considered objective function, that usually is the global maintenance cost, for the eventual constraints and for the resolution approach [4], [5], [6], [7], [8].

Nevertheless, for some systems the failure event can be dangerous, too expensive or even disastrous. For these reasons, being the reliability defines as the probability that the system operates without failure for a fixed period of time under stated conditions, a high reliability level is imposed. That being so, the reliability constitutes another meaningful parameter to assess the system performance [9].

This paper tackles the problem of singling out the maintenance actions to execute on a series-parallel system exploiting planned plant stops, for example due to production change, so as a given reliability value is warranted up to the next planned stop.

Reference [10] tackles the problem of identifying the set of elements on which to operate during a planned downtime between two missions, aiming to maximize the system reliability up to the next mission. The maintenance activities must be completed within a stated time and a fixed cost. This decision-making process is referred to as “selective maintenance”. The problem is formulated by a mathematical programming model and two numerical examples with 10 and 12 components respectively are reported. Cost or time minimization are also considered as constraints together with the reliability.

* Corresponding author: Tel.: (0039) 091-23861826; Fax: (0039) 091-7099973; E-mail: acerta@dtpm.unipa.it
Reference [11] deals with the reliability maximization for a system constituted by a series arrangement of subsystems, each one containing a set of identical parallel arranged elements. A constraint on maintenance time is considered. Elements have a constant failure rate and therefore maintenance action reduces to the replacement of those failed. The decision variable is the number of failed elements to be replaced for each subsystem. The problem is a non-linear knapsack problem and the Authors propose four improvements to speed up the total enumeration approach originally employed [12].

When lots of failure data are available for some components or they are monitored, the failure probability estimation can be considered reliable. On the contrary, for other components, estimates could be the more vague the longer their operating time is. For the decision maker it is important to take into account such vagueness in order to single out the best solution. That solution must also be robust, that is scarcely sensitive to possible differences between the supposed component reliability and the real one. Therefore, the decision maker could be interested in having more optimal solutions for the problem previously formulated, obtained by assuming, at least for few components, different reliability values inside some vagueness range. In this way, he/she can verify how much robust a solution is in relation to various possible scenarios. This analysis could drive the decision maker in choosing a solution characterized by a lower level of reliability than the optimal one obtainable under the hypothesis of sure reliability data, but more robust in that case a state of uncertainty exists.

In the present paper, it is supposed to handle with uncertain reliability data. Consequently, the decision maker has to evaluate the most robust solution for different possible scenarios. To this purpose, a parameter is proposed for representing the solution robustness. Moreover, an exact algorithm is presented for selecting the system elements to be maintained. Such algorithm is an adaptation of a previous one designed by the same authors [13], of which a short description is given in section 3.

In the next section, the faced problem is justified and it is analytically expressed. After the resolution algorithm presentation, the fourth paragraph proposes a parameter for the evaluation of solution robustness. Lastly, a numerical example explains the whole procedure and final remarks conclude the paper.

2. PROBLEM FORMULATION

The system taken into consideration is a productive system operating for process. The plant is periodically stopped for production change and system setup. During these stops, maintenance actions are carried out in order to maximize the system reliability. As a matter of fact, for this type of system, failure is an event that must be avoided because it is too expensive and/or too dangerous for workers. A time constraint is also introduced: maintenance must take place during the planned plant stop to execute system setup. This constraint, regarding the maximum maintenance time length, arises from the will of maximizing the system availability. This maximization finally constitutes a further object.

System reliability up to the end of the next utilization period can be calculated by the reliability values of its components. These values depend on the execution or not of maintenance actions during the considered plant stop. They can be regarded as sure values for some elements, in particular those ones that have cumulated a short use time or have been maintained. As a matter of fact, the mission time is assumed to be short as to the element life and the maintenance action is a “as good as new” type. Other elements characterized by sure reliability values are those working under monitored conditions. On the contrary, the reliability of other elements is affected with uncertainty. In particular, it is supposed that the real value of reliability of each of these elements falls into a range that can be estimated. The uncertainty arises from a poor knowledge of the actual system operative conditions and then it can be supposed that a high correlation exists among the reliability values affected with uncertainty. Consequently, in any case, all the real reliability values are in the same position of their vagueness range.

Each maintenance action involves both a resource engagement time and a spare part cost. If more maintenance crews can simultaneously operate and the interventions do not present precedence constraints, then the global time required to carry out the interventions singled out to be optimal is given by:

\[
T = \sum_{i \in I} t_i / n
\]

where \(I\) is the set of elements selected to be maintained, \(t_i\) is the maintenance time on the element \(i\) (\(i=1, ..., N\)) and \(n\) is the number of crews.

The previous problem can be formulated as follows:
where \( R \) is the system reliability and \( T^* \) is the planned downtime length for the system setup.

In order to calculate the system reliability it is necessary to know the reliability values for all constituting elements. As some values are doubtful, a definite value must be fixed inside each uncertainty range. To this aim, each continuous range is substituted by \( S \) equidistant values and then each element is characterized by \( S \) different scenarios. As said before, concerning the reliability, it can retain that the same conditions occur for all elements and so the system will operate in a scenario in which all elements have a reliability value given by their own first scenario, or by the second, … or by the \( S\)th scenario. To sum up, \( S \) scenarios need to be considered and, for each of them, the elements to be maintained have to be individuated by solving the previously formulated problem. Anyway this problem is a NP-hard combinatorial problem [12] and, even if it could be easily expressed in terms of mathematical programming, the presence of both Boolean variables and a non-linear objective function makes this approach the more difficult the bigger the problem dimension is.

3. THE PROPOSED ALGORITHM

Consider a system constituted by series components, some of them, in their turn, constituted by elements in parallel-series disposition.

Regarding to the series components, constituted by only one element, they are ordered in a list however drawn out. The algorithm matches the two possible maintenance states of the first element with the two possible maintenance states of the second one. Among the four obtained sequences, the algorithm eliminates those dominated, if they exist. A sequence \( s_1 \) dominates a sequence \( s_2 \) if:

\[
R_{s_1} > R_{s_2} \quad \text{and} \quad T_{s_1} \leq T_{s_2}
\]

or

\[
R_{s_1} = R_{s_2} \quad \text{and} \quad T_{s_1} < T_{s_2}
\]

The survived sequences are matched with the two possible states of the next element in the list. The procedure continues until the last element is considered.

At each step, in order to reduce the number of partial sequences to be considered at the next step, two further elimination criteria can be introduced. In fact, each partial sequence \( s \) evolves to a maintenance time included between two extreme values. The first one, a Lower Bound value, \( LB_T(s) \), is obtained if no maintenance is executed on the remaining elements. The other, an Upper Bound value \( UB_T(s) \), is obtained hypothesizing that all the remaining components are maintained. Comparing these two time values with the constraint value, two cutting criteria are defined. LB criterion: if \( LB_T(s) > T^* \), then the partial sequence \( s \), even if non dominated, can be eliminated because it can not respect the constraint (3). UB criterion: if a partial sequence verifies the condition \( UB_T(s) \leq T^* \), then all the others having a lower reliability are removed. Moreover, the survived one does not require to be further branched and it is completed by maintaining all the remaining components.

For each component constituted by elements parallel-series, each branch is considered separately and it is analyzed in the same way of the series systems. That is, the algorithm preliminarily eliminates the dominated sequences of a branch, saving those non dominated. The parallel is subsequently solved by considering all the possible combinations of the survived sequences of the first branch with those ones of the second branch and, as before, only the non dominated partial sequences are saved. The method iteratively continues by adding the non dominated sequences of the next branch to those saved at the previous step.

Firstly the algorithm solves the components constituted by the parallel-series elements, obtaining for each of them the non dominated sequences, and later it analyzes the overall system in which the parallel-series components will be considered as a series component characterized by the survived sequences, representing alternative states,
rather than the two states maintenance yes or not.

4. MEASURE OF THE ROBUSTENESS

After solving the problem expressed by equations (1) to (3) for all considered scenarios, the following quantities can be calculated:

- $R(i)$, system reliability with respect to the optimal solution obtained for the scenario $i$;
- $R(j|i)$, system reliability when the implemented solution is that one determined for the scenario $i$ but that occurring is $j$.

Then, $L(j|i) = R(j) - R(j|i)$ measures the loss of reliability when the scenario $j$ happens while the scenario $i$ is erroneously considered. An expected loss value can be evaluated by:

$$L(i) = \sum_{j=1}^{S} L(j|i) / S$$

(6)

It is obvious that such summation should be minimized and, at the same time, it is opportune to relate it to the reliability value of the selected solutions. After all, the robustness of a solution determined for the specific scenario $i$ can be defined by using the following ratio:

$$\text{Rob}(i) = R(i) / L(i)$$

(7)

that takes into account both the good quality of a solution, by means of the related reliability, and the possible reliability loss if a different scenario from that assumed occurs.

Finally, after the optimal solution has been found out for each hypothetic scenario and values of robustness have been calculated, the solution to be selected will be that one having the maximum value of robustness.

5. EXAMPLE

The procedure has been applied to a simulated example involving a system constituted by 44 components in series. Five components are macro-component constituted by parallel-series elements, while the others are constituted by just one element. If the macro-components are codified in the form $j(k; m_1, \ldots, m_r, \ldots, m_k)$, where $j$ is the generic component, $k$ indicates the number of branches and $m_r$ the elements in series in the branch $r$, then the encoding of the macro-components is:

1(3; 6,6,6) ; 2(2; 4,4) ; 3(2; 1,1) ; 4(3; 3,3,3) ; 5(2; 2,2)

so the total number of elements that constitute the system is eighty. The input data are reported in Table 1. $R_{i,b}$ is the reliability ($10^8$) of a generic element $i$ at the end of the next mission if it is not maintained, $R_{i,a}$ is the reliability under the hypothesis of maintenance and $t_i$ is its execution time. If two values are reported for $R_{i,b}$, then the element reliability is doubtful: the two values represent the uncertainty range and they coincide with the values assigned to the extreme scenarios 1 and 5.

In the example the uncertainty range has been further divided into ten scenarios ($S=10$). The maintenance time constraint has been changed acting on the parameter $w$ in the relation $T^* = wT_{\text{max}}$, being $w$ a positive number smaller than 1 and

$$T_{\text{max}} = \frac{\sum_{i=1}^{N} t_i}{n}$$

(8)

the time required to maintain all system elements.

Figures 1a and 1b show, just for example, the reliability values that allow to calculate the robustness of the different solutions obtained by setting the parameter $w$ equal to 0.5. The curve $R(i)$ is the same in both figures and it
Robust Maintenance Scheduling for Reliability Maximization

is relevant to the reliability values of the optimal solutions obtained varying the scenario. Curves \( R(j|i) \) express the reliability that the solution individuated for the supposed scenario (1, 10 or 5) should have under the hypothesis of realization of the scenario \( j \).

Table 1: Input data

<table>
<thead>
<tr>
<th></th>
<th>( R_{i,b} )</th>
<th>( R_{i,a} )</th>
<th>( t_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91930489</td>
<td>99009072</td>
<td>12.9</td>
</tr>
<tr>
<td>2</td>
<td>91916844</td>
<td>99017004</td>
<td>12.0</td>
</tr>
<tr>
<td>3</td>
<td>91924971</td>
<td>99016256</td>
<td>11.3</td>
</tr>
<tr>
<td>4</td>
<td>91924935</td>
<td>99015988</td>
<td>8.5</td>
</tr>
<tr>
<td>5</td>
<td>91926427</td>
<td>99017716</td>
<td>10.5</td>
</tr>
<tr>
<td>6</td>
<td>91923929</td>
<td>99016700</td>
<td>9.5</td>
</tr>
<tr>
<td>7</td>
<td>91923803</td>
<td>99017301</td>
<td>6.9</td>
</tr>
<tr>
<td>8</td>
<td>91930728</td>
<td>99016298</td>
<td>8.8</td>
</tr>
<tr>
<td>9</td>
<td>91925076</td>
<td>99013058</td>
<td>10.1</td>
</tr>
<tr>
<td>10</td>
<td>91922546</td>
<td>99015311</td>
<td>11.6</td>
</tr>
<tr>
<td>11</td>
<td>91933422</td>
<td>99018656</td>
<td>7.2</td>
</tr>
<tr>
<td>12</td>
<td>91922546</td>
<td>99013912</td>
<td>12.1</td>
</tr>
<tr>
<td>13</td>
<td>91931273</td>
<td>99018048</td>
<td>6.1</td>
</tr>
<tr>
<td>14</td>
<td>91915307</td>
<td>99020536</td>
<td>10.5</td>
</tr>
<tr>
<td>15</td>
<td>91924484</td>
<td>99020339</td>
<td>12.5</td>
</tr>
<tr>
<td>16</td>
<td>91915365</td>
<td>99014839</td>
<td>12.1</td>
</tr>
<tr>
<td>17</td>
<td>91928219</td>
<td>99017534</td>
<td>10.7</td>
</tr>
<tr>
<td>18</td>
<td>95797552</td>
<td>99626280</td>
<td>9.9</td>
</tr>
<tr>
<td>19</td>
<td>95787842</td>
<td>99631849</td>
<td>9.6</td>
</tr>
<tr>
<td>20</td>
<td>95795488</td>
<td>99631002</td>
<td>12.4</td>
</tr>
<tr>
<td>21</td>
<td>95790784</td>
<td>99625609</td>
<td>7.4</td>
</tr>
<tr>
<td>22</td>
<td>95794939</td>
<td>99625553</td>
<td>10.5</td>
</tr>
<tr>
<td>23</td>
<td>95794122</td>
<td>99631586</td>
<td>6.9</td>
</tr>
<tr>
<td>24</td>
<td>95793123</td>
<td>99624232</td>
<td>11.1</td>
</tr>
<tr>
<td>25</td>
<td>95792950</td>
<td>99628841</td>
<td>6.5</td>
</tr>
<tr>
<td>26</td>
<td>83396499</td>
<td>98435609</td>
<td>12.6</td>
</tr>
<tr>
<td>27</td>
<td>85118183</td>
<td>98436647</td>
<td>12.8</td>
</tr>
<tr>
<td>28</td>
<td>84149720</td>
<td>98075810</td>
<td>7.7</td>
</tr>
<tr>
<td>29</td>
<td>84181065</td>
<td>98077443</td>
<td>7.8</td>
</tr>
<tr>
<td>30</td>
<td>84185946</td>
<td>98079900</td>
<td>12.4</td>
</tr>
<tr>
<td>31</td>
<td>8415043</td>
<td>98073956</td>
<td>8.5</td>
</tr>
<tr>
<td>32</td>
<td>84165735</td>
<td>98084021</td>
<td>5.6</td>
</tr>
<tr>
<td>33</td>
<td>84165296</td>
<td>98078055</td>
<td>12.1</td>
</tr>
<tr>
<td>34</td>
<td>84188378</td>
<td>98080317</td>
<td>11.7</td>
</tr>
<tr>
<td>35</td>
<td>91587808</td>
<td>99315673</td>
<td>7.9</td>
</tr>
<tr>
<td>36</td>
<td>91528130</td>
<td>99313566</td>
<td>11.3</td>
</tr>
<tr>
<td>37</td>
<td>92515189</td>
<td>99310754</td>
<td>8.4</td>
</tr>
</tbody>
</table>
The reliability values graphically represented in these figures allow to evaluate the solution robustness for the scenarios 1, 5 and 10. Figure 2 shows the robustness values for all scenarios.

The solution to be selected is that having the greatest robustness, hence that one obtained for the scenario $s=7$. Anyway, the scenario corresponding to the greatest robustness requires some consideration.

When uncertain data are available, but precise values must be utilized in order to optimize some objective function, the prudence could induce to assume mean values inside the variability ranges: this choice ought to minimize the opportunity loss. That seems to be confirmed by curves in Fig. 1: the scenario $s=5$ gives rise to the lowest losses of reliability, but losses for $s=7$ are less meaningful as to the respective reliability value and then the solution obtained for $s=7$ is more robust. Curves in Figure 3, that concern various constraints about the available maintenance time, confirm that a prudent choice is not always the best: if the most robust solution is obtained with $s=5$ when $w=0.1$ or $w=0.9$, the best solution is obtained with the scenario $s=6$ if $w=0.05$ and with $s=8$ if $w=0.7$. 
After all, it can be stated that prudence does not pay. In the same way, it can be affirmed that either optimism or pessimism do not pay: see the robustness falls for scenarios with high indexes when $w=0.7$ and $w=0.9$, and the low values obtained for $s=1$ and $s=2$ when $w=0.5$. The tuned procedure is able of determining the best solution whatever the decision maker mood is.

Figure 4 points out that the most robust solutions are obtained when $T^*$ is very low or approaching $T_{\text{max}}$.

This fact can be justified by the following considerations.

By varying the considered scenario, the optimal solutions are generally different, hence reliability losses, $L$, (see equation (6)) occur. Whenever $w$ takes a high (low) value, the set of elements $I$ that can be maintained within $T^*$ is very numerous (very scanty), consequently the optimal solutions for the different scenarios are very similar among them since the eventual differences are restricted to the few elements excluded from (included in) $I$. If $w$ approaches 1 (approaches 0), all elements (no elements) are maintained and all solutions are identical whatever scenario may
come true: the robustness goes to infinity. On the contrary, if $w$ assumes an intermediate value then the optimal solutions variability with relation to the scenario is high. These conditions bring to a low robustness value.

Lastly, about the efficiency of the proposed optimization procedure, the global run time, from the determination of the optimum solution for all the scenarios up to the singling out of the most robust one for given value of $w$, requires less than 15 seconds.

6. CONCLUSIONS

The present paper has been focused on a system reliability maximization problem by maintaining the components during planned intervals of given length. Since some components reliability values can be affected by vagueness within a given range, the proposed procedure allows singling out a solution that is not only optimal in a classical point of view but also robust, i.e. not very sensitive to changing conditions. A suitable parameter is proposed for measuring such robustness which calculation requires that the maximization problem is solved several times. This is possible thanks to a previously developed algorithm allowing to solve such problem to optimality and in a very short time, even for complex series-parallel systems.

REFERENCES


A model for a periodic preventive maintenance policy for a series-parallel system

A. Certa, G. Galante, T. Lupo*, A. Passannanti and G. Passannanti

Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale
Università degli Studi di Palermo,
Palermo, 90128, Italy

ABSTRACT

Aim of the present paper is to develop an optimal periodic preventive maintenance policy for a stochastically deteriorating system. The problem concerns both the determination of the optimal number of system stops in a finite horizon and the components set on which to perform the maintenance actions with relation to each planned stop. The stop frequency is investigated among a set of discrete values because, for many production plants, the preventive maintenance actions can not be carried out at any time. The maintenance actions have to assure a very high reliability level between two consecutive stops at the minimum global maintenance cost. This aspect is meaningful for those systems whose stops due to a failure may imply not only a considerable loss of production and repair costs, but also high risks for safety of people and/or environment. The maintenance cost includes that directly related to the element maintained, but also the system downtime cost and the inspection cost. The problem is formulated for a series-parallel system by a non-linear integer mathematical programming model and solved by varying some meaningful parameters.

1. INTRODUCTION

The main part of problems on preventive maintenance (PM) for multi-component systems [1] [2] already tackled in literature concern the determination of the optimal interval after which performing the maintenance actions on same components or on previously determined group of components. The last problem is defined as block replacement problem [3].

The most treated goal in PM problems addressed to single out the PM intervals is the global maintenance cost minimization while the most used parameter in evaluating the production system performance is the stationary availability, also defined as the expected percentage of time in which the system is working. For example, Bris et al. [4] developed a single objective optimization model, with periodically inspected and maintained components, aiming to optimize the maintenance policy for every single component system by minimizing the cost function for a given availability constraint. In order to solve the problem, the Authors propose a genetic algorithm, whose structure includes the first inspection time and the interval length between two maintenance interventions for each component. Tsai et al. [5] developed an algorithm to individuate a PM policy for a multi-component system. For each component the PM interval is investigated on the base of the availability maximization and the minimum intervals value is chosen as the system PM interval. Caldeira Duarte et al. [6] proposed a mathematical programming model for a series system to calculate the interval between the PM actions for each component, by minimizing the maintenance cost and guaranteeing a required level of system availability.

Since for production plants unexpected interruptions generally imply high penalty costs, the PM is required to reduce the failure probability, that is to improve the reliability level defined as the probability that a system will operate without failure for a stated period of time under specified conditions. Moreover, for some continuous operating systems, like chemical processing facilities, power plants, aircrafts, ships, etc., the failure event can be not only expensive but also dangerous or even disastrous. For these reasons, a high reliability level is required during the operating period. Then the reliability system, besides the total maintenance costs, is here chosen like a driver parameter to individuate an optimal PM policy for multi-component system.

Tam et al. [7] proposed a PM model for a multi-component system, based on the imperfect maintenance concept, that simultaneously optimises both costs and reliability. The Authors divide the system life cycle into equal planning periods within which the model plans the possible action (maintaining the component, replacement of the

* Corresponding author: Tel.: (0039) 091-23861879; Fax: (0039) 091-7099973; E-mail: lupo@dtpm.unipa.it
component or doing nothing) to be performed with relation to each component such that the net present value is minimized and a minimum reliability level is assured. The model, applied to a numerical example with thirty periods and three components, is solved by an evolutionary algorithm. The same Authors [8] proposed three simple models to determine the optimal maintenance intervals for problems with different managerial requirements: minimum required reliability, maximum allowable budget and minimum total cost. Since in product system the maintenance actions can not be performed at any time, but strictly within a stated downtime, a more realistic model is proposed forcing the maintenance interval to be a multiple of a priori given value. The optimization approach consists in two steps. In the first one, using the Excel Solver tool, the cost function is minimized. In the second step, by an exhaustive analysis, the maintenance intervals are arranged so as the interventions occur at given instants. The exhaustive analysis imposes that the procedure can be applied if the number of components is low. Moreover, the Authors considerably simplify the reliability system evaluation by hypothesizing that the generic component reliability at any time is given by the reliability value at the end of its optimal maintenance interval. Lapa et al. [9] developed a genetic algorithm to evaluate the PM policy by means of cost and reliability parameters. The algorithm also allows the flexible intervals use between maintenance interventions.

The model here proposed deals with the following constrained problem: given a system whose components can be maintained only during the planned stops, the aim is to individuate the optimal number of system stops in a finite horizon and, for each planned stop, to single out the set of elements that have to be maintained so as a required reliability level is warranted up to the next stop with a minimum total maintenance cost.

To our knowledge the problem herein considered is not widely handled in literature. Cassady et al. [10] tackled this type of problem with relation to a single stop. The purpose is to individuate the set of elements on which to operate in a planned break between two missions in order to maximize the system reliability for the next one. Maintenance activities must be completed within an allotted time and cost. The problem is formulated by a mathematical programming model and solved for two simple examples. Certa et al. [11] handled the problem related to only one stop, developing a model that provides the system functioning with a required reliability until the next planned stop and with the minimum maintenance cost. The downtime is required as short as possible and in any case not larger than the planned stop. The model is solved by a genetic algorithm and a mathematical programming approach.

The present paper is organized as follow: the problem is formulated in section 2, the proposed mathematical model is reported in section 3. Section 4 presents the reliability calculation and section 5 shows the results related to a case study.

2. PROBLEM FORMULATION

The considered system consists of several series components, some of which constituted by parallel-series elements. It is supposed that the maintenance actions, also on the parallel elements, must be carried out within the planned system stop making the generic element as good as new. The maintenance plan is related to a finite horizon \( H \) within which the number of system stops has to be individuated. The elements on which to perform maintenance actions have to be identified at each stop, during which inspections on parallel elements have also to be carried out. For many production plants the maintenance actions can not be done at any time, that is there are some operating constraints forcing to chose the stop frequency among some discrete values. Therefore, it is supposed that the optimal number of planned stops belongs to a discrete set.

The time between two consecutive stops is coinciding with operating time \( t_p \) because the latter is very longer than the maintenance time. Moreover, \( t_p \) is equal to \( H \) divided by the number of stops.

As before said, for some continuous operating systems, the stop for failure implies a high production loss and maintenance costs as well as risks for workers. For these reasons, the failure probability has to be lower than an accepted value, that is it is needed to choose the elements on which to perform the maintenance actions so that the system reliability between the restart and the next planned stop is not lower than a fixed value. If a system failure occurs, a minimal repair is performed on the failed element, so the maintenance actions imply a “good as old but unbroken” state for the element. This assumption involves that the age of the failed elements is not reset to zero after a failure repair and it maintains its value. The hypothesis of minimal repair is acceptable because the guaranteed reliability level during the system operating time is very high, that is the failure probability is very low. For the same reason the corrective maintenance cost is negligible and thus it is not taken into account.

For the fixed period taken into account, the model here proposed aims to minimize the global maintenance cost, that is the sum of costs at each stop. For each planned stop the maintenance cost is composed by three terms. The
A model for a periodic preventive maintenance policy for a series-parallel system

first one is directly related to the generic element and it includes the spare part cost, necessary to re-establish the
normal operating conditions of the element, and the maintenance crew cost. Since the downtime implies a loss of
production, the second term of cost is the downtime cost, supposed to be directly proportional to the longest time
among the maintenance time of each element maintained during the stop. This assumption arises from the
hypothesis that the maintenance crews operate in parallel because the downtime cost is supposed to be very high.
The last term represents a constant inspection cost occurring at any planned stop.

3. MATHEMATICAL PROGRAMMING MODEL

The following nomenclature is hereafter used:

- $H$: considered finite temporal horizon;
- $i=1...N$: index representing a generic element;
- $l=1...L$: index representing a generic series component;
- $j=1...Ps$: index representing a generic planned stop;
- $k=0...Ps$: index representing the last planned stop in which maintenance action has been performed for the
  value $1...Ps$ and represents the start of the system for the value 0;
- $s$: index representing a generic scenario related to different stop numbers over $H$;
- $x_{i,j}$: binary variable taking 1 if a maintenance action is performed on the element $i$ at the planned system
  stop $j$;
- $z_{i,j,k}$: binary variable taking 1 if, with relation to the planned stop $j$, the last instant of maintenance action
  on the element $i$ has been performed at the planned stop $k$;
- $r_{i,j,k}$: reliability of the generic element $i$ calculated at a generic planned stop $j$, for the next stop $j+1$, when
  the last maintenance action has been performed at the stop $k$;
- $c_i$: maintenance cost related to a generic element $i$;
- $t_i$: time to perform the maintenance actions on the element $i$;
- $t_p$: system operating time between two consecutive planned stops;
- $T_j$: time to perform the maintenance actions at a generic planned stop $j$;
- $T_{start}$: the age of the system at the planned stop $j=1$;
- $C_i$: global maintenance cost over $H$;
- $C_{adT}$: unit time cost of system downtime;
- $C_c$: inspection cost for the planned system stop;
- $R_{i,j}$: reliability of the element $i$ calculated at a generic stop $j$, for the next stop $j+1$;
- $R_{Sys,j}$: system reliability calculated at a generic stop $j$, for the next stop $j+1$;
- $R^\ast_{Sys}$: minimum accepted value of system reliability.

The problem described in the previous section can be formulated as an integer non-linear mathematical
programming model:

\[
\min C_s \quad \forall s
\]  

with

\[
C_s = \sum_{j=1}^{Ps} \sum_{i=1}^{N} c_i \cdot X_{i,j} + C_{adT} \cdot \sum_{j=1}^{Ps} T_j + P_s C_c
\]
Subject to:

\[ R_{\text{Syn},j} \geq R_{\text{Syn}}^* \quad \forall j \]  

As before mentioned, the considered system consists of several components in series, some of them constituted by parallel-series elements. System reliability, for each planned stop, can be expressed in terms of components reliability by:

\[ R_{\text{Syn},j} = \prod_{i=1}^{l} R_{i,j} \]  

The reliability of the component \( l \) is calculated on the base of the reliability \( R_{i,j} \) of each element belonging to it and on the base of the reliability relations among them.

At a generic planned system stop the reliability of the element \( i \) is expressed as below:

\[ R_{i,j} = r_{i,j} \cdot x_{i,j} + \sum_{k=0}^{l-1} (r_{i,j,k} \cdot z_{i,j,k}) \cdot (1 - x_{i,j}) \]  

Whenever \( x_{i,j} \) is equal to 0, for each element the reliability is a conditional probability whose calculation is explained in the next section.

With relation to the previous constraint, it is necessary to introduce another constraint that links the \( x_{i,j} \) value to the \( z_{i,j,k} \):

\[ z_{i,j,k} = x_{i,k} \cdot \prod_{n=k+1}^{l} (1 - x_{i,n}) \quad \forall k < j \]  

As said in section 2, for each planned stop the system downtime is equal to the maximum time value among those necessary to perform the maintenance actions on selected elements. This assumption is assured by imposing the following constraint:

\[ T_j \geq t \cdot x_{i,j} \quad \forall j, \forall i \]  

because the equation (2) has to be minimized.

By solving the model for different stop numbers, that one assuring the minimum total maintenance cost will be chosen for the PM policy.

4. RELIABILITY CALCULATION

It is assumed that the system is composed by stochastically deteriorating elements with a Weibull life time distribution. Moreover the PM actions lead the elements to the “as good as new” state.

For the generic element \( i \) the reliability function \( R_i(t) \) at the time \( t \) is given by the following expression:

\[ R_i(t) = \exp \left[ -\left( \frac{t}{\eta_i} \right)^{\beta_i} \right] \]  

where; \( \beta_i > 1 \) is the shape parameter and \( \eta_i > 0 \) is the scale parameter of the distribution.

Referring to a generic system stop \( j \), the reliability of the element \( i \) at the next planned system stop \( j+1 \) depends on the instant of the last performed maintenance action on the considered element. If the last maintenance action is performed at the stop \( j \), the reliability \( r_{i,j} \) can be calculated by using the expression (9):
A model for a periodic preventive maintenance policy for a series-parallel system

\[ r_{i,j} = \exp \left[ \frac{-\left( t_p \right)^\beta}{\eta_i} \right] \quad \forall j \]  

(9)

where \( t_p \) denotes the system operating time between two consecutive planned stops.

Instead, if the last maintenance action on the considered element is not performed at the stop \( j \) but it was performed at the stop \( k \), with \( k < j \), the reliability \( r_{i,j,k} \) is a conditioned probability, that is the probability that the element properly operates at stop \( j+1 \) given the working condition at the stop \( j \) and with the “as good as new” state at stop \( k \).

According to the theory of probability, such conditioned probability \( r_{i,j,k} \) is given by the following equations:

\[ r_{i,j,k} = \frac{R_{i,j} \left( t_{j+1} - t_k \right)}{R_{i,j} \left( t_j - t_k \right)} = \frac{R_{i,j} \left( (j+1-k) \cdot t_p \right)}{R_{i,j} \left( (j-k) \cdot t_p \right)} \quad \forall j, \forall k \mid k \neq 0 \]  

(10)

\[ r_{i,j,k} = \frac{R_{i,j} \left( t_{j+1} \right)}{R_{i,j} \left( t_j \right)} = \frac{R_{i,j} \left( j \cdot t_p + T_{\text{start}} \right)}{R_{i,j} \left( (j-1) \cdot t_p + T_{\text{start}} \right)} \quad \forall j \text{ and with } k = 0 \]  

(11)

where \( T_{\text{start}} \) is the age of the system at the stop \( j=1 \).

Finally, using the Weibull distribution the relations (10) and (11) become:

\[ r_{i,j,k} = \exp \left[ \frac{-\left( (j+1-k) \cdot t_p \right)^\beta}{\eta_i} \right] \quad \forall j, \forall k \mid k \neq 0 \]  

(12)

\[ r_{i,j,k} = \exp \left[ \frac{-\left( (j-k) \cdot t_p \right)^\beta}{\eta_i} \right] \quad \forall j \text{ and with } k = 0 \]  

(13)

5. ILLUSTRATIVE EXAMPLE

The proposed model is applied with relation to a medium case study for the series-parallel system described in Fig.1.
As said in the previous section for each element it is hypothesized a Weibull lifetime distribution whose parameters $\eta$ and $\beta$ are reported in Table 1. The table also shows the costs data, expressed in monetary units (M.U.) and time data in time units (T.U.). It is supposed that horizon $H$ is 2400 T.U. and the operating time between two consecutive stops can assume 480, 600, 800, 1200 T.U., that is four scenarios have to be considered related to $P_s$ equal to 5, 4, 3 and 2. At the beginning of the period $H$ the reliability of the system is 0.992 which is the value corresponding after about $T_{start}=1500$ T.U. of continuous system running. The scenario with $P_s=1$ has not been considered because it does not assure the constraint on $R_{Syst}^*$. 

The model has been solved by the commercial software Lingo and the table 2 shows the elements selected for the maintenance interventions with relation to $C_{uDT}=15$ M.U., $C_c=45$ M.U. and $R_{Syst}^*=0.995$ for the different scenarios. It is possible to note as the examined scenario $s=4$, related to two stops and $t_p=1200$ T.U. is the optimal PM planning. The scenarios $s=1$ and $s=2$ are characterized by 2 stops in which not even one element is individuated, thus only inspection activities are carried out.

Solutions for each scenario strongly depend on the initial system conditions, the value of $C_{uDT}$ and the minimum accepted value of system reliability. In fact, by varying only one of these parameters the solution considerably changes. In table 3 the solutions obtained by varying $C_{uDT}$, $R_{Syst}$ for the scenario 2 are reported. In the scenario with $C_{uDT}=1$ and $R_{Syst}=0.9965$ maintenance interventions are individuated for each planned stop. In fact, the downtime cost is considerably lower than the other cost terms and thus it is convenient to take advantage of all the planned stops to make maintenance actions.

![Considered system](image.png)
A model for a periodic preventive maintenance policy for a series-parallel system

6. CONCLUSIONS

This paper develops a PM model that can be applied to plan maintenance actions for continuous industrial system, production line system or military system that can only operate with a very high reliability level. In fact, the model represents a valid support tool to the decision maker when the failure event implies high risks for workers and
unacceptable costs. It allows to establish a maintenance plan over the entire horizon, based on several strategic parameters as reliability and costs, making easier the management of activities as supplying of the materials, resource allocation, etc. In particular, the problem here tackled pertains to the simultaneously individuation of the optimal number of planned stops in a finite horizon and the selection of the elements to be maintained. The problem is formulated as a mathematical programming model with the aim of minimizing the global maintenance cost at the same time assuring a certain system reliability level over the finite horizon. Future development can regard a multi-objective formulation using a model in which different stakeholders, as outsourcer and maintenance service provider, have conflicting interests.

REFERENCES


A Multi-Objective Fleet Routing Optimization for Passenger Shipping among Islands

L. Cannizzaro, M. Enca, C. M. La Fata and G. Passannanti

Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale
Università di Palermo
Palermo, 90128, ITALY

ABSTRACT

In the last few years the statistical analysis on maritime freight and passenger transport highlights a positive trend of demand and for the next future, given the stressing congestion characterizing the road transport, a further increase both on the deep-sea-shipping and the short-sea-shipping is predicted. Given that, the navigation companies are called to improve their own services to assure a high quality and efficient customers transport, deriving at the same time their own benefits by minimizing the operating costs. Hence, an efficient and careful routes planning activity plays a meaningful role on navigation company productivity and income. Furthermore, the fleet deployment also constitutes a company critical aspect from both the management and maintenance system point of view. Starting from a real case of maritime hydrofoil passengers transport, the tackled problem consists in determining the optimal fleet deployment and the minimum cost routes of a priori known fast vessels fleet in order to satisfy the passengers requests. The passengers splitting is here allowed and, also, vessels are permitted to perform sub-cycles, therefore each port can be reached more than once even by the same vessel. An integer nonlinear multi-objective mathematical programming model is formulated for the fleet deployment and routing design and solved by the Goal Programming Method in order to previously find the minimum size of fleet, successively, the optimal routes and finally minimizing the passengers ride-time. A case study is solved by a commercial software and the results are reported.

1. INTRODUCTION

In the last few years the statistical analysis on maritime freight and passenger transport highlights a positive trend of demand and for the next future, given the stressing congestion characterizing the road transport, a further increase both on the deep-sea-shipping and the short-sea-shipping is predicted. Given that, the navigation companies are called to improve their own services to assure a high quality and efficient customers transport, deriving at the same time their own benefits by minimizing the operating costs. Anyway, solving real transport problems is a complex task especially if carried out without any technological support and knowledge about the specific scientific literature. Since an efficient routes planning plays a meaningful role on navigation company productivity and income, one of the most challenging issue on transport problems is the routing design; in fact even a modest routing improvement can result in large monetary savings. Nowadays the main part of literature on ship routing pertains to industrial operations while references on tramp and liner shipping are fewer. Since liner operate according to a published itinerary and schedule, similar to a bus line [1], maritime passengers transportation which the present paper is focused on can be assimilated to the liner shipping. Anyway, taking into account the evident differences as the uncertainty due the sea-weather conditions, Christiansen and Fagerholt [2] highlight that much of research within the general transport routing problems can be adjusted to solve the ship routing and emphasize the meaningful differences between vehicle and ship planning problems. Particularly, the authors show how these differences can be taken into account when searching for efficient routes and schedules and propose a real planning problem within short-sea-shipping similar to a multi Pick-up and Delivery Problem with Time Window (m-PDPTW) pointing out the relative few contributions existing for ship routing design. An interesting real liner shipping problem is presented by Fagerholt [3] in order of deciding the optimal weekly routes for a given fleet of ships operating on a hub and spoke network in which all spokes nodes have a transportation request to the hub. Firstly, all feasible routes, in terms of respect of time and capacity constraints for each ship, are generated together

* Corresponding author: Tel.: (0039) 091-23861826; Fax: (0039) 091-7099973; E-mail: lafata@ditra.unipa.it
with the cost and the duration for each route by solving a Multi-trips Vehicle Routing Problem (m-VRP), next the
routes are given as input to an integer programming (IP) problem solving which routes for each ship are selected
such that total transportation costs are minimized and each port demand is satisfied. A Travelling Salesman Problem
with Allocation, Time Window and Precedence Constraints (TSP-ATWPC) for a real bulk ship routing problem is
instead considered and solved by a dynamic programming algorithm by Christiansen and Fagerholt [4]. The problem
is a combined multi-ship, pick-up and delivery problem with time windows and multi-allocation problem,
thus allowing multiple products to be carried simultaneously by the same ship. A meaningful mixed-integer
nonlinear model for finding a minimum cost routing for a heterogeneous fleet of ships engaged in pick-up and
delivery liquid bulk products among islands is formulated by Al-Khayyal and Hwang [5]. Products to be handled
are assumed to require dedicated compartments in the ship while the problem is finally to decide how much of each
product should be carried by each ship from supply ports to demand ports. Concerning the passengers shipping
routing, to our knowledge only few researches have been conducted up to now. An interesting work focused on
modeling the operation of passenger vessels that proposed by Giziakis et al. [6]. The authors approach the problem
by a simple linear programming problem on a network of 37 nodes in the Aegean, considering two different kinds of vessels with different speed and capacity and disregarding of demand fluctuations.

Even the fleet deployment constitutes a company critical aspect from both the management and maintenance
system point of view. Several researches have been developed as regard to fleet deployment in order to optimize the
management and the assignment politics. An optimal fleet deployment approach for the feeder system is proposed
by Fagerholt [7] that is also interested in designing coherent weekly routes for ships in the fleet. The paper
addresses of deciding the minimum cost fleet in a real liner shipping problem taking into account that each
product should be carried by each ship from supply ports to demand ports. Concerning the passengers shipping
determination of vessel traffic under annual known supply and demand constraints where total fuel costs and port
dues are minimized. The estimation of fleet size is carried out in a constrained network of current ship routes.
Subsequently, the problem is analyzed on a full relaxed network where all routes are allowed by formulating the
problem as one of vehicle routing problem (VRP) nature.

Starting from a real case of maritime hydrofoil passengers transport, the problem here presented consists in
determining the minimum cost routes and the corresponding assignment of a priori known fast vessels fleet in order
to satisfy the passengers requests. Each pick-up node is characterized by several transport demands having different
drop-off demand estimations. Furthermore, the passengers and their passenger splitting is allowed only once per week.
Therefore, each port can be reached more than once even by the same vessel. The just mentioned scenario and the corresponding hypothesis perfectly match the reality in which the navigation companies operate.
To our knowledge the considered problem is not yet tackled in the literature; furthermore it strongly differs from the
classical Pick-up and Delivery Problem (PDP) or Dial-A-Ride Problem (DARP) in which every available vehicle is
forced to travel a single trip, visiting just once nodes with a non-zero demand, and consequently splitting using the
same vehicle are not taken into account. The reminder of the present paper is organized as follow: the problem and
the proposed mathematical model are defined in Section 2, computational results on the real case study are detailed
in Section 3, while final conclusions are drawn in Section 4.

2. MODEL DESCRIPTION

A fleet deployment and routing design problem is here proposed in order to get the minimum fleet use and the
minimum cost routes so that all passengers requests are fulfilled. Forward, every trip departing from the node fixed
as hub, visiting nodes, included the hub, even more than once, and returning to the hub is defined route. Hence,
vessels are permitted to perform sub-cycles and passengers splitting is allowed even by the same vessel. In this
paper, the search for the optimal solution is carried out in three steps by the Goal Programming Method [9] concomitantly optimizing three different objective functions: firstly, the fleet deployment is minimized in order to
single out the minimum number of vessels needed to perform the service; next, the minimum path is found and
lastly the total number of departing passengers is minimized in order to force all drop-off operations to be properly
carried out. Let \( G = (V, A) \) be an undirected and symmetric graph wherein \( V = \{1, \ldots, Q\} \) is the set of nodes (seaports) to be visited and \( A \) is the arc set between all nodes. The hub will be hereafter denoted by node 1. For each pair of
locations \((i,j) \in A\), \(j \in V \) the nautical miles travel distance \( d_{ij} \) of going from \( i \) to \( j \) and the demand \( pax_{ij} \) associated to
each node \( i \) and directed to \( j \) are given. Finally, let \( K = \{1, \ldots, Q\} \) be the fleet vessels set and \( C_i \) is the capacity of the

\[ G=\left(\mathbb{V},\mathbb{A}\right) \]
generic vessel $k \in K$. Finally, let $n = [1, \ldots, N]$ be the index indicating in which sequence nodes are visited through the route sailed by the vessel $k$. That being so, the following variables are introduced:

- $z_{kinj}$ binary variable taking value 1 if the vessel $k$, visiting the node $i$ as $n^{th}$ through its route, directly goes from $i$ to $j$, that is a direct arc between $i$ and $j$ is sailed by the specific hydrofoil $k$.
- $pax_{car}^{kinj}$ integer variable representing the number of passengers picked-up by vessel $k$ at node $i$ and headed to $j$, being $i$ visited as $n^{th}$ node through the sequence of vessel $k$. Obviously, $j$ must be at least the $(n+1)^{th}$ node during the same sequence.
- $pax_{scar}^{kinj}$ integer variable representing the number of passengers dropped-off at node $i$ and deriving from node $j$, being $i$ visited as $n^{th}$ node through the sequence of vessel $k$. Obviously $j$ must be at least the $(n-1)^{th}$ node during the same sequence.
- $pax_{nod}^{kin}$ integer variable representing the number of passengers to tally leaving the node $i$ visited as $n^{th}$ during the sequence of vessel $k$. This variable takes into account the passengers picked-up at $i$ and those already on board coming from the $(n-1)$ nodes previously visited by the vessel $k$ and having different destination.

### 2.1. Step 1: Fleet Deployment Minimization

The first objective function $F_1$ aiming to the fleet deployment minimization is finally formulated by the following equation:

$$
F_1 = \sum_{k} \sum_{j|i \neq 1} z_{kinj}
$$

subject to:

1. $\sum_{j|i \neq 1} z_{kinj} \leq 1 \quad \forall k \in K$ (2)
2. $\sum_{i} \sum_{n|i \neq 1} \sum_{j|i \neq 1} z_{kinj} \leq UP \cdot \sum_{j|i \neq 1} z_{kinj} \quad \forall k \in K$ (2a)
3. $LOW \cdot \sum_{j|i \neq 1} z_{kinj} \leq \sum_{i} \sum_{n|i \neq 1} \sum_{j|i \neq 1} z_{kinj} \quad \forall k \in K$ (2b)
4. $\sum_{k} \sum_{j|i \neq 1} \sum_{n} z_{kinj} \geq \text{stop} \quad \forall j \in V \mid j \neq 1$ (3)
5. $\sum_{i} \sum_{n} \sum_{j|i \neq 1} z_{kinj} \cdot d_{ij} \leq D_{k\text{ max}} \quad \forall k \in K$ (4)
6. $\sum_{j|i \neq 1} z_{kinj} - \sum_{j|i \neq 1} z_{k(n-1)i} = 0 \quad \forall k \in K, \forall n > 1$ (5)
7. $\sum_{j|i \neq 1} z_{kinj} = 0 \quad \forall k \in K, \forall i \neq 1, \forall n > 1$ (6)
\[
\sum_{n} pax_{\text{car}_{knj}} - \sum_{m \geq n(n+1)} pax_{\text{scar}_{jmi}} = 0 \quad \forall k \in K, \forall (i, j) \in A | i \neq j (7)
\]

\[
[pax_{\text{modo}_{k1}} - \sum_{j \neq 1} pax_{\text{car}_{k11j}}] \cdot \sum_{j \neq 1} z_{k11j} \geq 0 \quad \forall k \in K (7a)
\]

\[
pax_{\text{mode}_{k1}} = 0 \quad \forall k \in K, \forall i \in V | i \neq 1 (7b)
\]

\[
pax_{\text{scar}_{k1j}} = 0 \quad \forall k \in K, \forall (i, j) \in A | i \neq j (7c)
\]

\[
\sum_{k} \sum_{n} pax_{\text{car}_{knj}} = pax_{ij} \quad \forall (i, j) \in A | i \neq j (7d)
\]

\[
pax_{\text{modo}_{kn}} \leq C_k \cdot \sum_{j \neq i} z_{knj} \quad \forall k \in K, \forall i \in V, \forall n (7e)
\]

\[
pax_{\text{car}_{kn}} \leq C_k \cdot \sum_{r \neq j} \sum_{q \neq n} z_{knr} \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n (7f)
\]

\[
pax_{\text{car}_{kn}} \leq \text{UP} \cdot \sum_{r \neq j} \sum_{q \neq n} z_{knr} \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n (7g)
\]

\[
pax_{\text{car}_{kn}} = \sum_{m \geq n} pax_{\text{scar}_{jmi}} \leq 0 \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n (7h)
\]

\[
pax_{\text{scar}_{kn}} - \sum_{m \geq n(n-1)} pax_{\text{scar}_{jmi}} \leq 0 \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n > 1 (7i)
\]

\[
[pax_{\text{modo}_{kn(n-1)}} - \sum_{r \neq i} pax_{\text{scar}_{knr}} + \sum_{r \neq j} pax_{\text{car}_{knr}} - pax_{\text{modo}_{kn}}] \cdot z_{kn(n-1)} \leq 0
\]

\[
\forall k \in K, \forall (i, j) \in A | i \neq j, \forall n > 1 (7l)
\]

\[
z_{knj} + z_{kn(n+1)} + z_{kn(n+2)} + z_{kn(n+3)} \leq 3 \quad \forall (i, j) \in A | i \neq j (8)
\]

\[
z_{knj} \in [0,1] \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n (9)
\]

\[
pax_{\text{car}_{kn}}, pax_{\text{scar}_{kn}} \in Z^+ \quad \forall k \in K, \forall (i, j) \in A | i \neq j, \forall n (10)
\]

The objective function (1) aims to minimize the total number of hydrofoils departures from the hub seaport. Constraint (2) ensures that each departing vessel \(k\) leaves its initial position, while (2a) and (2b) guarantee that no arcs are sailed by vessel \(k\) if it does not leave the hub. \(LOW\) and \(UP\) are respectively the lower and the upper bound
on total number of arcs to be sailed by $k$ before definitively returning to node 1. In particular, $LOW$ is stated equal to 1, while $UP$ is declared a large number. For each node, constraint (3) forces the number of stops to be greater than a fixed value and (4) guarantees that each vessel sails a distance at most equal to $D_{\text{max}}$, being the latter the maximum distance allowed for vessel $k$ within the planning period. Reminding that during each route the generic vessel $k$ is allowed to visit the hub, but it is also forced to definitively return to it at the end of the planning period, (5) is the precedence constraint for the node 1, while the expression (6) is the flow balance constraint for all nodes unlike the hub. Constraints (7) to (7l) concern the passenger request satisfaction in regard to vessel capacity limit. In particular, for each arc and vessel, (7) imposes the total number of passengers picked-up at $i$ to be equal to the total number of passengers dropped-off at $j$ on condition that $j$ is visited after $i$; (7a) defines the passengers leaving the hub with regard to the first departure. (7b) and (7c) force $pax_{\text{node}_{k}i}$ and $pax_{\text{car}_{k}i}$ to be equal to zero with regard to the first departure. (7d) ensures the demand and satisfaction, while (7e) guarantees that the number of passengers leaving the specific node $i$ by vessel $k$ does not exceed the capacity. Furthermore, the latter constraint assures that $pax_{\text{node}_{k}i}$ is equal to zero whenever no arcs come out of $i$. Expression (7f) asserts that $pax_{\text{car}_{k}i}$ on vessel $k$ stopping over at $i$ does not exceed the capacity and, furthermore, it forces $pax_{\text{car}_{k}i}$ to zero whenever no arcs come out of $i$. (7g) forces $pax_{\text{car}_{k}i}$ to be equal to zero whenever no arcs enter on $j$. (7h) links the number of passengers picked-up by $k$ stopping over at $i$ with regard to stop $n$ to be less or at most equal to the total passengers successively dropped-off at $j$. (7i) links the number of passengers dropped-off by $k$ stopping over at $i$ with regard to stop $n$ to be less or at most equal to the total passengers previously picked-up at $j$. For each vessel $k$, (7l) defines passengers leaving the generic node $i$ with regard to the stop $n$. (8) avoids that vessel consecutively sails backward and forward the same arc. (9) and (10) define the variables $z_{kij}$, $pax_{\text{car}_{k}i}$ and $pax_{\text{car}_{k}j}$ to be respectively binary and positive integer. Hereafter, the optimal value of $F_1$ is indicated as $F_1^*$.  

2.2. STEP 2: ROUTING OPTIMIZATION

After solving the step 1, the second objective function $F_2$ to minimize aims to find the minimum distance routes and it is formulated as below:

$$F_2 = \sum_k \sum_i \sum_{j \neq i} \sum_{n} d_{ij} \cdot z_{kij} \quad (1)$$

subject to constraints (2)-(10) and to the following constraint concerning the first goal optimization:

$$F_1 \leq F_1^* \quad (12)$$

Hereafter, the optimal value of $F_2$ is indicated as $F_2^*$.  

2.3. STEP 2: PASSENGERS RIDE TIME OPTIMIZATION

The last optimization step, constrained to $F_1^*$ and $F_2^*$, concerns the departing passengers minimization in order to force all drop-off operations to be carried out in correspondence to the first destination stop after the pick-up.

$$F_3 = \sum_k \sum_i \sum_{n} pax_{\text{node}_{k}i} \quad (1)$$

subject to constraints (2)-(10), (12) and to the following constraint concerning the second goal optimization:

$$F_2 \leq F_2^* \quad (14)$$

2.4. NONLINEAR CONSTRAINTS LINEARIZATION

Since nonlinear constraints are time-consuming, starting from a proposition already mentioned by Al-Khayyal and Hwang [5] the (7a) and (7l) have been linearized as follow:
$$pax_{\text{-}\text{nodo}}_{k1} = \sum_{j:j\neq 1} pax_{\text{-}\text{car}}_{k1j} - C_k \cdot \sum_{j:j\neq 1} z_{k1j} \geq -C_k \quad \forall k \in K \quad (7a')$$

$$pax_{\text{-}\text{nodo}}_{kj(n-1)} = \sum_{r|r\neq j} pax_{\text{-}\text{car}}_{kjr} + \sum_{r|r\neq j} pax_{\text{-}\text{car}}_{jnr} - pax_{\text{-}\text{nodo}}_{kjr} + C_k \cdot z_{kj(n-1)} \leq C_k \quad \forall k \in K, \forall (i, j) \in A \mid i \neq j, \forall n > 1 \quad (7l')$$

that together with the (7c) assure the respect of conditions.

3. NUMERICAL CASE: THE ISOLE EGADI ARCHIPELAGO

The proposed mathematical programming model is applied to the real case of the Ustica Lines S.p.A., operating in the Mediterranean Sea (Sicilian Islands) for the fast vessels passenger shipping since 1993. With relation to a one day planning period, a network of 4 seaports is considered. In order to take into account the daily demand fluctuations, the planning period is furthermore divided into 4 time slots. The daily node-node passengers requests $pax_{ij}$ (Table 3) are calculated as average on historical data, hence the planning problem here tackled is defined as static. The nautical-miles distances $d_{ij}$ among nodes are reported in Table 4. In this context, a homogeneous fleet of two 364 seats capacity hydrofoils is considered, while the maximum distance allowed for each vessel is calculated taking into account the mean value of cruising speed for this kind of vessels. Finally, the commercial software LINGO is used to solve the model. The obtained solutions are reported in Table 1 and 2.

<table>
<thead>
<tr>
<th>Table 1: Solutions of time slot 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME SLOT</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>$F_1$</td>
</tr>
<tr>
<td>$F_2$</td>
</tr>
<tr>
<td>$F_3$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{nodo}}(1, 1, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{nodo}}(1, 1, 4)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{nodo}}(1, 2, 2)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{nodo}}(1, 2, 7)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{nodo}}(1, 3, 3)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 1, 1, 2)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 1, 1, 3)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 1, 4, 2)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 2, 2, 3)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 2, 7, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 3, 3, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{car}}(1, 3, 3, 2)$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 1, 4, 3)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 1, 8, 2)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 2, 2, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 2, 5, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 2, 5, 3)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 3, 3, 1)$</td>
</tr>
<tr>
<td>$pax_{\text{-}\text{scar}}(1, 3, 3, 2)$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 2: Solutions of time slot 3 and 4

<table>
<thead>
<tr>
<th>Time slot 3</th>
<th>Time slot 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROUTE</td>
<td>ROUTE</td>
</tr>
<tr>
<td>1–3–2–1–2–4–2–1</td>
<td>1–3–2–4–2–3–1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROUTE</th>
<th>F₁</th>
<th>F₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>1032</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROUTE</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>2</td>
<td>127</td>
<td>136</td>
</tr>
<tr>
<td>3</td>
<td>364</td>
<td>360</td>
</tr>
<tr>
<td>4</td>
<td>103</td>
<td>157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROUTE</th>
<th>F₅</th>
<th>F₆</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>2</td>
<td>127</td>
<td>136</td>
</tr>
<tr>
<td>3</td>
<td>364</td>
<td>360</td>
</tr>
<tr>
<td>4</td>
<td>103</td>
<td>157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROUTE</th>
<th>F₇</th>
<th>F₈</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>2</td>
<td>127</td>
<td>136</td>
</tr>
<tr>
<td>3</td>
<td>364</td>
<td>360</td>
</tr>
<tr>
<td>4</td>
<td>103</td>
<td>157</td>
</tr>
</tbody>
</table>

Table 3: Passengers request $p\text{ax}_{ij}$

<table>
<thead>
<tr>
<th>Time slot</th>
<th>Node 1</th>
<th>Node 2</th>
<th>Node 3</th>
<th>Node 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 585</td>
<td>59 0</td>
<td>0 266</td>
<td>29 51</td>
</tr>
<tr>
<td>2</td>
<td>136 0</td>
<td>1 9 0</td>
<td>131 0</td>
<td>8 13</td>
</tr>
<tr>
<td>3</td>
<td>41 5 0</td>
<td>0 4 11</td>
<td>0 4 71</td>
<td>32 0 0</td>
</tr>
<tr>
<td>4</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 29 27</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Table 4: Distances node–node $d_{ij}$

<table>
<thead>
<tr>
<th>Node</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>9</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>9</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3</td>
<td>-</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>10</td>
<td>13</td>
<td>-</td>
</tr>
</tbody>
</table>

For each time slot planning, the Figure 1 shows the obtained solutions. As regards to slots 1 and 3, arcs are characterized by a letter to clarify the vessel sequence.
4. CONCLUSIONS

A nonlinear multi-objective mathematical programming model for the fleet deployment and routing design is proposed and solved for a real case of maritime hydrofoil passengers transport. To our knowledge, the passengers shipping has not been widely treated in literature; furthermore the problem here tackled strongly differs from the classical PDP or DARP since vessels are permitted to perform sub-cycles and, also, the passengers splitting even by the same vessel is allowed. In our future work, it will look at extending the current model to a larger case by using a heuristic approach for routes generation together with the mathematical programming.

ACKNOWLEDGMENTS

It is expressed a particular appreciation to the Ustica Lines president Mr. Vittorio Morace.

REFERENCES

AIDC Technologies and their impact on Management Information Systems

Athanasios Klonis1* and Farhad Nabhani1

1School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

Information Technology (IT) and Information Systems gather, process, store and disseminate information across the manufacturing support systems. Despite the high degree of sophistication of such software, the basic underlying principle of any information processing activity remains; the output is proportional to the input. The complexities in business environments inherently result in data inaccuracies. This inaccurate data causes expensive manual interventions, which add to the non-value added activities and impact on cost. The use of Automatic Identification and Data Collection (AIDC) technologies has been known to be a powerful tool in achieving higher data accuracy and increasing the quality of information. The paper describes the significance of information in manufacturing and manufacturing support systems and provides an overview of the different AIDC methods and relative applications.

1. INTRODUCTION

Manufacturing has traditionally been described as the process of transforming raw material into finished products. As such, the manufacturing process is evident from ancient times to modern civilisation. Our ancestors realised that it was possible to create goods for the exchange of food, and the idea of exchange commerce was born as an alternative to the existing agricultural activities. Skills and techniques were developed to transform raw material into useful objects, signalling the beginning of the era of craftsmen. Developments and innovations in technology as a result of the first industrial revolution changed the manufacturing environment. The innovations in process technology, the development of factories and the introduction of machines allowed for the growth of industrial units and gave birth to the production system, where people, equipment and procedures were organised to fulfil the manufacturing operations of an organisation [1]. The incursion of IT in modern production systems may be distinguished in the automation of the physical production through the utilisation of numerically controlled machines and machine centres, robotics and Flexible Manufacturing Systems (FMS) and automation of the production support systems through information processing technologies [2]. Production support systems do not interact directly with the product; however they are responsible for its movement and progression through the factory and include the set of procedures, data and knowledge used by the company to manage production and logistics to successfully deliver a product to the end customer [3]. Management Information Systems (MIS) include any hardware and software which collect, compile, store, analyse and distribute information across the organisation’s activities [4]. The first software development for planning and scheduling materials for complex manufactured products was Materials Requirements Planning (MRP). The software aimed to reduce the amount of capital spent by a manufacturing organisation on raw materials and Work-In-Progress (WIP) inventory by effectively managing Bills of Materials (BOM), inventory, work orders and production scheduling on a large scale [5]. These calculations and MRP’s effective operation, relied on input data such as the Master Production Schedule (MPS), BOM, inventory records and lead times.

With the evolution of technology and the developments of low cost micro-processing, these systems found widespread adoption as they were proven to improve the decision making process [6]. Despite the undisputable success of MRP, the system was limited to the material handling in the organisation and failed to incorporate other paramount functions in production such as human resources, machines and capital [7]. This led to the development of its successor in the 1980s, known as Manufacturing Resource Planning or MRP II. The aim of MRP II was to

* Corresponding author: Tel.: +44 1642 342482; Fax: +44 1642 342401; E-mail: t.klonis@tees.ac.uk
fully integrate and synchronize all the resources within an organization by using the MPS to ensure the most cost effective manufacturing method. MRP II included virtually any resource entering into production, including human resources, machines, material and capital. By linking all activities to the MPS, the organization efficiency was increased by removing unnecessary jobs, removing delays due to late deliveries and reducing stocks of WIP because of the smooth material flow through the full manufacturing process. MRP II is often referred to as a “closed loop system” because the production functions automatically feedback to the master production schedule, and the input is adjusted to modify the system [7].

Improving efficiency in the production processes was not enough for organisations to remain competitive. Their supply chain network had to be improved and the goal of eliminating waste throughout the activities of the organisation expanded in its transport and distribution networks. The concept of logistics was born with the primary aim to deliver the right product, in the right time at the right place while reducing lead time from order to delivery through appropriate coordination of material and information flow [7]. Developments in information technology such as the internet, Electronic Data Interchange (EDI) and telecommunications provided organisations with the opportunity to develop real-time linkages across the supply chain [8]. Improved and faster communication of information through the exchange of Point of Sale (POS) information with suppliers and replenishment on actual sales reconfigured the supply chain [9]. Supply Chain Management and EDI, set the foundation for the development and rapid expansion of multi-module application software integrating all the departments of an enterprise, known as Enterprise Resource Planning (ERP) [7]. ERP systems first appeared in the early 1990s and initially provided inventory control [5]. The system expanded to include MRP and MRP II, and finally integrated all departments and functions such as sales and order management, marketing, purchasing, warehouse management, financial accounting and human resource management into a single computerised system to serve all the particular needs for delivering a product or service [5].

Whilst MRP implementations gained widespread adoption in the western world, the Japanese developed a replenishment system which ensured that parts were made in response to a demand [10]. The system is known as Kanban from the Japanese word for signalling card. Kanban has been described as “an information system which harmoniously controlled the production quantities in every process” [10]. Kanban is highly responsive to actual customer demand; however it is also overwhelmed by non-value adding activities. Kanbans are best suited to parts of repetitive production demand and also a fairly level schedule [11]. In the event of spikes in demand, Kanban sizes for each purchase part number must be recalculated manually. This type of recalculation rarely takes place in environments that have thousands of part numbers resulting to significant buffer inventory and shortages. Additionally, every triggered Kanban has to be communicated manually via fax or phone to the supplier. The reliance on the human element to activate the replenishment trigger and the communication flow between the requisitioner and the supplier could result in shortages as Kanbans could be misplaced. Although, Toyota’s Kanban system was successful without relying on IT, the technological advances in the field of Information Technology (IT) may reduce the non-value adding activities associated with manual Kanban systems while creating a highly responsive replenishment system. It has been suggested that a Kanban system integrated to MRP can benefit from the automatic recalculation of Kanban part numbers every time MRP is generated [11]. Additionally through EDI, the system may communicate a projected demand to the supplier each time MRP is run which may be used for capacity planning while automatically creating a purchase order to the supplier, when a supplier Kanban item is triggered by consumption [11]. Finally, through the application of AIDC technologies incoming bar-coded Kanban signals can be used to automatically bring up purchase orders and determine container quantities [11].

2. DATA, INFORMATION AND AIDC TECHNOLOGIES

Although similar, information and data are in fact two distinct entities. Data is the “raw material” that is input to be processed and information is the “product” or output of the process. This is illustrated schematically in Figure 1. From the illustration it is clear that the output depends on the input. A key for the success of IT based systems has always been the uploading with accurate and timely data and several authors have stressed the importance of quality data input in the computing world with the phrase “Garbage in, Garbage Out” [4]. Data collection refers to the capture of data items concerning an event, at the place where the event occurs and at the time of its occurrence [12]. Data entry is the subsequent process in which the collected data is input into a computer. Traditional manual data collection and data entry have been proven insufficient, time consuming and prone to errors due to the human intervention [4, 12, 13].
Automatic Identification and Data Capture (AIDC) technologies have evolved to reduce or even eliminate errors in data collection and manual data entry and to serve a number of processes in numerous industries such as logistics, retail, and manufacturing [14]. These systems were primarily developed due to the need to collect data in a faster and more efficient manner [12]. AIDC technologies may be distinguished in feature extraction and data carrier technologies [15]. Feature extraction AIDC technologies exploit object, action or property relations and include biometrics and machine vision, which measure physical and behavioural characteristics [15]. Popular examples of feature extraction techniques include fingerprints, hand geometry, iris and retina scanning, facial recognition, voice verification and signature verification [16]. Data carrier AIDC technologies include barcode, optical character recognition, magnetic stripes, smart cards, memory cards and radio frequency tags [12, 17].

2.1. LINEAR BARCODES

Barcodes have been used successfully in industry for more than 50 years and are by far the most widely used AIDC technology. They are found in literally every product that is commercially available today, mainly due to the ease of implementation and low installation costs. Barcode technology has also found applications in manufacturing support systems such as material control in warehouses, tool control, shipping and transportation. Linear bar codes are parallel arrangements of width or height modulated bars and spaces, which encode data in only one direction according to a distinct pattern known as the symbology [18]. An alternative way of encoding linear barcodes is the bumpy or 3D barcode, which is encoded using smooth and bumpy areas rather than light and dark spaces. In width modulated barcodes the data is encoded by varying or modulating the widths of the individual bars and spaces. In contrast, in height modulated symbologies the data is encoded by varying the heights of the individual bars. The ultimate dimensions of the barcode symbol are determined by the amount of information to be contained on the code [13]. The two types of data encoding are illustrated in Figure 2.

![Width vs. Height Modulated barcodes](image)

Traditional barcodes are vertically redundant, meaning that the same information is repeated vertically. This allows the symbol to be scanned despite the presence of defects within a certain area in the barcode. As the bar height increases, so does the probability of at least one path being readable (See Figure 3). The height of the code does not add any additional data capacity when utilizing a width-modulated linear barcode symbology. A linear bar code symbol typically consists of five parts: a quiet zone, a start character, data characters, a stop character, and a final quiet zone [12, 18]. Linear barcode systems reliability is dependent on the quality of the produced barcodes,
which in turn is affected by the medium on which the barcode symbol is printed as well as the production method employed.

Figure 3: 1D barcode

2.2. 2D BARCODES

Traceability of parts is paramount in numerous industries including the automotive, aerospace and pharmaceutical industry, as it is an essential element of quality assurance in the manufacturing process [19]. Despite the success of linear barcodes in retail applications, limitations of the technology in terms of data capacity and size have rendered it inadequate for traceability applications requiring encoding of significant amounts of data in many size-limiting environments [20]. Manufactured components are required to hold large volumes of data to provide the necessary data for keeping track of which parts have been fitted to which assembly, which supplier supplied them and which batch they came from [19]. These reasons have led to the development and widespread use of 2D barcodes. Similar to the linear barcode symbologies distinction in width and height modulated, 2D symbologies are distinguished in stacked (multi-row) and matrix (area) symbologies [18]. This distinction is made on the way that data is encoded and decoded. 2D barcodes encode data vertically as well as horizontally. Since both dimensions contain information, at least some of the vertical redundancy is gone. Therefore 2D barcodes utilize an error correction algorithm which allows for the recovery of information even when labels are damaged. However the error correction algorithm encodes the same data in a significantly larger area as illustrated in Figure 4.

Figure 4: Stacked Symbologies

2D stacked bar code symbologies consist of multiple rows of width modulated bars and spaces, where each row has the same physical length and looks like a conventional linear symbology [18]. Data Matrix is the widest used matrix symbology, mainly because its availability in the public domain [21]. The code’s high data capacity offers increased traceability as due to its small size it may be applied to the smallest of components. Data matrix is an alphanumeric, variable length symbology with advanced error correction capabilities. Data Matrix may be encoded in a variety of area sizes ranging from 0.5mm² to 144 mm² and can encode up to 3116 numeric characters [21]. The ultimate size of the code relies upon the type and volume of data to be encoded, the substrate, the production method and the capability of the reading equipment [22].

Production of data matrix is preferable via Direct Part Marking machines which permanently mark the code on the metal surface. The use of the latter result in a deep mark on the component’s surface that is more likely to survive surface corrosion [19]. The code may be successfully read even when part of the code is damaged, missing or
AIDC Technologies and their impact on Management Information Systems

obscured as the correct answer can be deduced from the data values immediately before and after [21]. Data Matrix is recommended by organisations representing the automotive, aerospace and electronic manufacturing industries such as the Automotive Industry Action Group (AIAG), the Air Transport Association (ATA) the International Aerospace Quality Group (IAQG) and the Electronic Industries Alliance (EIA) due to the permanence of the mark as well as its error correction abilities [22].

2.3. OPTICAL CHARACTER RECOGNITION

The identification of printed text by a computer, aided by a scanning device, is known as Optical Character Recognition (OCR). The OCR uses a number of well-defined fonts, OCR-A, comprising of 0 to 9, A to Z, and a few special characters and the OCR-B. The two most common fonts OCR-A and OCR-B are shown in Figure 5.

![OCR fonts](image)

Figure 5: Optical Character Recognition fonts

OCR fonts can be printed using different techniques and are readable by humans as well as machines. OCR is a two-dimensional technology, as an OCR scanner examines both vertical and horizontal features of the printed characters during the decoding process. Developments in the font-independent scanning equipment have been made possible by advances in recognition algorithms and the decreasing cost of computing, but this technology has not yet had much impact on traditional automatic identification applications. This technology is primarily used as an entry technique for word processing and desktop publishing systems [18].

2.4. MAGNETIC STRIPES

This is the second most popular keyless identification method after barcodes [4, 13, 18]. This method encodes data onto a magnetic stripe which may be applied on any non-magnetic material. The magnetic stripe is made from many small magnetic particles bound together in a resin [12]. Typical application examples include credit cards, identification cards and railway tickets [14]. Within the magnetic stripe there are three tracks within which data may be encoded as illustrated in Figure 6. Their characteristics, functional usage and relative position have been defined by the British Standard Institute in BS EN ISO/IEC 7811-2:1996.

![Magnetic stripe](image)

Figure 6: Location of magnetic stripe tracks (52, 53)
According to the standards tracks one and three are defined as 210 bits per inch and track two as 75 bits per inch. The bit density in conjunction with the data format dictates how much data is encoded on each track. Track one may encode up to 79 characters, track two up to 40 characters and track three up to 107 [12]. Magnetic stripes are distinguished by their ability to withstand corruption by other magnetic fields, known as coercivity. High Coercivity (HiCo) stripes are normally black and Low Coercivity (LoCo) stripes are a lighter brown. The fractional difference in price has established the use of the black stripes, which are more durable and may encode more data. Magnetic stripe readers do not differentiate as to whether a stripe is HiCo or LoCo and are capable to read both. A typical magnetic stripe automatic identification system consists of a magnetic stripe encoding device (normally a printer), a magnetic stripe reader (or magnetic head) and the stripe medium. Because of their magnetic nature, magnetic stripe cards should be kept away from magnetic fields, as this may cause loss of data.

2.5 SMART CARDS – INTEGRATED CIRCUIT CARDS

A smart card is a portable device used for automatically identifying items, transactions or events, and which can become an integral part of an event [12]. All smart cards have the same basic components: memory for storage of data, memory for the chip operating system (COS), an integrated circuit chip which acts as a central processing unit (CPU) and read/write interface. More advanced devices incorporate communications software as well as encryption algorithms to make the application software and data unreadable. Data is stored in the chip’s memory and can be accessed to complete various processing applications. When used in conjunction with the appropriate applications, smart cards can provide enhanced security and the ability to record, store, and update data. Some of the current applications of smart cards include bank and telephone cards, secure access control and in-home entertainment.

2.6 RFID

RFID is the abbreviation for radio frequency identification. RFID is an automatic identification and data collection technology that uses radio waves to identify people or objects carrying encoded microchips. The technology is the product of military research which began in the early 1940’s [23]. Modern applications of RFID may be found in remote controls for garage doors, access control systems (see for example [24]), anti-theft devices for the automotive industry (immobiliser), new European union passports, libraries (see for example [25]), education (see for example [26]) and agriculture (see for example [27]). The simplest form of RFID is the 1-bit transponder which is the technology used in Electronic Article Surveillance (EAS) Systems. Electronic Article Surveillance was developed in the 1960’s to ensure the security and safety of nuclear materials [28]. Since then, the technology has found widespread use in the retail industry as it has been found to be more effective than other crime prevention methods such as security guards or store redesign [28].

In general, all RFID systems consist of a tag (or transponder) and a reader as shown in Figure 7. The reader is responsible for activating the transponder by creating a magnetic field and emitting an interrogation signal which then modulates to read and write data on the tag [29]. The power source provides electrical power to the other tag elements. It is also used as a method for classifying RFID tags. Therefore, the presence of an internal power source that is used to activate the tag distinguishes the tag as active whereas the absence of a power source distinguishes the tag as passive. Active transponders rely on a dedicated power supply such as a battery. Passive tags do not have a dedicated power supply but are activated from the incoming radio frequency signal. Because they do not have a dedicated power source, passive tags are considerably smaller than active tags but operate at a limited distance, normally a few centimetres to a metre. On contrast active tags may be read at longer distances and at higher speeds. LF transponders use induction coils while UHF and Microwave frequency tags use antennas.

RFID tags are available in four frequencies, namely Low Frequency (LF) at 120-140 kHz, High Frequency (HF) at 13.56 MHz, Ultra High Frequency (UHF) at 433, 860 – 960 MHz and Microwave or super high frequency at 2.45 GHz and above [29]. LF and HF tags are generally passive, activated by near field inductive coupling and are therefore suitable for applications requiring small amounts of data at relatively low speeds and at short distances, normally from 2 – 100 cm.
3. SUMMARY

Under constant cost pressures, organisations must relentlessly find ways and make use of appropriate technologies to reduce the non-value adding activities throughout the value stream to satisfy their end customers without having to outsource their operations. By definition, only material processing activities are value adding while information processing activities are necessary but non-value adding. They constitute a supportive function, necessary for the delivery of the product or service. Both sets of activities are of equal importance. For instance, without an appropriate information flow, earlier processes cannot make exactly what the later process requires. Organisations are unaware of what customers require, what inventory exists to meet those requirements and when more of a product should be produced or ordered. Accurate and immediate information in management information systems enable better decision making and is a necessity in management control as inaccurate data causes expensive manual interventions which in turn add to the non-value added activities and impact on cost.

The use of Automatic Identification and Data Collection (AIDC) technologies is a powerful tool in reducing unnecessary handling of information, achieving higher data accuracy and improving the quality of information. AIDC applications in industrial environments have been wide and varied; from simple keyboard replacements for automating data input, to assigning a product code for identifying and tracking an object as it is passed through various locations or processes during its lifecycle. AIDC can dramatically reduce transaction time, offer real time visibility of stock, help make more accurate production plans and improve the efficiency of ERP systems [30].

Increased traceability and improved inventory management across the supply chain have been recognised as a source of competitive advantage by world class organisation around the globe [31]. Traditional AIDC technologies are sometimes inadequate to provide the level of information required in current supply chains. Radio Frequency Identification (RFID), due to its physical characteristics and increased data capacity, is a promising solution for enabling more efficient supply chain collaboration [32] as each individual item can report its location across the supply chain [33].

REFERENCES


Comparison of two tools for the measurement of interfragmentary movement in femoral neck fractures stabilised by cannulated screws

Dr. Edward J. Bradley¹ and Prof Farhad Nabhani¹*

¹School School of Science & Technology
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

Achieving stability at the site of femoral neck fracture is an important factor for callus formation in the post-operative period. However, measuring interfragmentary movement in vivo is not currently possible as telemetric screws have not been manufactured for surgical use. Understanding how the implantation of the screws can affect the stability of the fracture allows the surgeon to tailor the procedure to the patient and produce the best possible outcome. Two techniques have been developed that measure interfragmentary movement between fractured surfaces. The first was a FEA model of the proximal femur with screws represented by nodal links. Movement was quantified by the amount of relative motion occurring between paired nodes either side of the fracture. The second was a mechanical compression test of a composite femur that allowed the motion analysis of paired markers on the external surface of the femur. Movement was digitised with markers selected and displacements calculated by transforming the global coordinate system to a local system relative to the fracture plane.

Validation of both techniques was achieved as inferior axial displacement of the femoral head was found to be similar to that recorded in the literature of between 0.6mm and 0.75mm. The maximum interfragmentary movement recorded using the FEA model was 0.0446mm and was 4.89mm using the motion analysis technique. The finite element model and the motion capture analysis techniques were found to both valid and repeatable tools for the measurement of the performance of cannulated screws for the fixation of femoral neck fractures. The advantage of finite element technique is the inclusion of realistic simulation of muscular and joint forces, which support the joint and result in the lower movements reported. However the motion capture analysis technique utilises real screws and enables the examination of screw thread dependent factors such as screw tightening levels.

1. INTRODUCTION

Fracture of the femoral neck is a common clinical condition [1], which has a high morbidity and mortality rate in the elderly [2-4]. In healthy patients with a long life expectancy and with the likelihood for the return to everyday activities, the most appropriate treatment for fractures of the femoral neck is minimally invasive surgery to reduce the fracture and provide adequate stabilization [5]. For undisplaced intracapsular fractures of the femoral head and neck, the use of cannulated hip screws to achieve adequate fixation is advocated [3 - 9]. Achieving immobilization of the fractured surfaces to allow adequate callus formation during the immediate post operative period is an important factor for satisfactory bone healing [10]. Instability at the fracture interface may inhibit the formation of callus, which characterises primary bone healing, and lead to fibrous tissue development and non-union [11]. Reduction of movement by compression of the fracture fragments have shown to improve callus formation [12, 13], while animal studies have shown that shear movement at diaphyseal fracture sites inhibited primary bone healing by damaging the external callus formation, disrupting the bone bridging process [52] and doubling the amount of fibrous tissue development [14].

* Corresponding author: Tel.: (01642) 342482; E-mail: f.nabhani@tees.ac.uk
It is therefore appropriate that surgical techniques are aimed at improving the stability, more importantly that they are able to decrease the inhibitory shear movement as this will increase the likelihood of healing failure.

The in vivo clinical measurement of stability is limited to radiological assessment of developing callus formation [14]. However, the movement occurring at the fracture interface when the bone is subjected to external forces cannot be measured in such a manner. It is necessary to use in vitro methods to determine the efficiency of fixation methods prior to applying to the human body or during development to ascertain if a new technique holds any benefit to the patient before it is used in surgery. Testing can be categorised into two distinct areas. The first is physical in vitro analysis, where the in vivo situation is replicated and loads are applied to the structure via a mechanical test machine. The second type of test is conducted using computer simulation. Finite element analysis currently is the tool of choice for this [15]. It is important to understand the advantages and limitations of these procedures to determine which, and when, is the most appropriate to employ.

Various test rigs have been designed to analyse the response of the proximal femur [16-19]. In the majority of these a single loading condition is simulated. These allow a dynamic linear load to be applied, with the orientation of the femur determining the type of loading condition. The most common type is an axial load to the femoral head. Such test rigs have been employed by Cristofolini et al [20] to assess the biomechanical behaviour of whole bone models, or to determine the mechanical strength and stability of fixated fractures of the femoral neck. Examples of the latter are the work by Selvan et al [5] in measuring the mechanical strength of different cannulated screw configurations, and that of Baitner et al [16] who examined the fracture gap increase at a single point, occurring in vertical fractures. Motion capture analysis is a technique that involves the sequential recording of location markers. This technique is widely used in gait analysis [21] or sports biomechanics [22, 23] to analyse gross human movement, such as running or tennis serve. In these situations relatively large movements are measured by capturing the position of markers applied to the body segment or object being analysed. By combining the experimental test rigs with the motion capture analysis system it is possible to create a tool for measuring interfragmentary movement.

It is important for understanding the femoral neck fracture fixation environment that it is tested under complex 3-dimensional loading regimes that occur in every day activities such as the gait cycle. Finite element analysis is a numerical mathematical method for solving engineering problems. It is possible to use this method to simulate the femur and apply the necessary complex loads. Finite element analysis has produced excellent results in identifying the stress distribution and load sharing patterns in surgical implants such as the retrograde intramedullary nail [24] where a removal of either a proximal or distal locking screw increases axial force by up to 17% and bending moments by 25%. In intramedullary nail [25] the introduction of a slot in a lag screw increases deformation and reduces stress, while the displacements in a trochanteric fracture were similar between a Gamma nail and a PFN [26]. The same finite element model of a femur with a Gamma nail implant produced greater deformation of a subtrochanteric fracture than a femoral neck fracture [27] and the deformation was reduced with the inclusion of an iliopsoas force. These studies highlight the strength of finite element analysis in producing a wide range of findings that can transferred to the clinical application of fixation techniques.

The aim of this study was to compare two in vitro techniques for measuring the level of interfragmentary movement in fixated femoral neck fractures. In each technique the advantages and limitations arising in the application of the technique will be highlighted.

**METHODS**

**FINITE ELEMENT METHOD**

An 8-bit, 3-dimensional voxel representation of the right-sided femur was created from CT images and using Visual Toolkit (public.kitware.com), an isosurface of the outer bone was obtained (threshold level = 71 for bone) and then input into MSc Mentat, the finite element pre-processor. The surface was meshed volumetrically using 188,882 tetrahedral elements (Fig. 1). Individual elements were assigned region-specific Young’s modulus based on local CT numbers using the following equation:

\[
E_{element} = E_{tissue}(GV_{element})^{\gamma}
\]

(1)

Where \(E_{element}\) is the value assigned Young’s modulus for each individual element. \(E_{tissue}\) is the tissue modulus (i.e. 5GPa). \(GV_{element}\) is the local relative grey-value set within the range between 0 and 1 and was calculated relative to the maximum and minimum CT numbers obtained for the entire object. A gamma value (Eq. 1) of 1.5 was assigned [28]. The main reason for using this approach is that it allows some representation of trabecular...
Figure 1. Finite element model of the proximal femur (viewed anteriorly). Different shades, as indicated by the scale, represent the region specific Young’s Modulus (GPa) applied to the model.

architecture even for low resolution CT images [28]. For validation purposes the model was subjected to a uniaxial load of 0.8kN. In order to compare the behaviour of this model with experimental data [20] the conditions of the laboratory experiment were recreated (Fig. 2A). Accordingly, the complete femur model was rotated to an 11° adducted position and subjected to an inferior directed load of 0.8kN. The point of load application was the most superior node on the femoral head.

Figure 2. A) The full femoral model used for validation. B) A modified model (Model A) used in subsequent analyses showing the fracture plane. Visible are the rigid links representing screws and loads applied to model (though loading shown is not indicative of any specific stage of loading). C) The fracture site as viewed from a medio-superior position.

A fracture (Fig. 2B) was simulated in the model at 50°. Contact elements were assigned to each side of the fracture site and a stick-slip model of frictional contact ($\mu = 0.5$) was used to represent the interfacial conditions for
bone-on-bone contact. In vivo joint contact force values recorded from various patients [29] were averaged to form a representative loading dataset and averaged forces were applied to the model at the femoral head. A variety of different muscle forces occurring during such activities have been predicted using modeling techniques [30]. The loads created by the abductors were included in the analysis, as applied by Polgar et al. [31]. as this muscle group, inserting at the greater trochanter, has been found to exert large forces during the early stages of the gait cycle. Based on a review of radiographic images taken from patients in the North-East of England [5], four models with a combination of three screws in a triangular configuration, and one model with two linearly positioned screws were created. The cross-sectional appearances of these models are shown in fig. 2C. Each screw is represented by attaching nodes on either side of the osteotomy and effectively attaching a node in the femoral head to a node in the lateral side of the femur with rigid links. These links were positioned to intersect the osteotomy to form the screw position patterns of Models A-E. As in the case of surgery, attempts were made to ensure that the screws were oriented approximately parallel to one another. In order to quantify this interfragmentary motion, paired nodes across the fracture site were identified and relative nodal positions were calculated. The individual interfragmentary displacements, representing axial and shear movement, was calculated from the difference in the relative position in each dimension.

MECHANICAL TEST PROCEDURE

Mechanical testing was performed on a 3rd generation composite femur (Pacific Research Laboratories Inc., Vashon Island, WV, USA) using a L6000R Universal Testing Machine (Lloyd Instruments, Fareham, UK) connected to a personal computer. The composite femur is composed of a solid polyurethane shell surrounding an E-glass-filled epoxy resin to accurately represent cortical and cancellous bone respectively. The distal femur was potted using polymethylmethacrylate (PMMA) (Sawbones, Sweden), covering the whole of the femoral condyles, at 11° adduction (Fig. 3) and the load was applied through a 1kN load cell to the superior region of the femoral head using a circular platen. Initially, the intact composite femur was tested to determine the overall validity of the experimental set-up, based on that of Cristofolini et al. [20]. An 800N load, representative of average body weight and load at mid-stance phase of the gait cycle, was applied to the femoral head at a loading rate of 80N·sec⁻¹. The peak axial deflection and load-displacement curve of the superior region of the femoral head was recorded. A total of 8 tests were conducted, with the whole setup being removed between each test. The force-displacement curve indicated that the synthetic bone displays linear properties.

Figure 3: Test rig set up showing whole bone composite femur at 11° adduction.
An osteotomy was created from the superior femoral neck to the inferior femoral neck so the osteotomy was angled 50° to the horizontal and the plane of the osteotomy was perpendicular to the longitudinal axis of the femoral neck. This osteotomy was designed to represent an undisplaced, Pauwels type II and Garden Stage II, fracture of the femoral neck. The femoral neck fractures created were reduced and stabilized with three large (6.5mm) cannulated hip screws (Synthes Ltd, Welwyn Garden City, UK) placed in a triangular form with a single superior apical screw and two in the inferior region [5]. The screws were tested at four levels of pretensioning, the first representing an untightened screw, with the screw inserted until the head came in contact with the external surface of the cortical bone. The second represented an initial tightening of the screw, with the screw turned a further half a revolution after the screw head came in contact with the external surface. The final two levels represented a single full revolution and two full revolutions of the screw.

A series of paired markers were placed around the osteotomy at 1cm intervals, with one marker on the femoral shaft side of the fracture interface and the corresponding marker on the adjacent femoral head side of the interface. Any relative movement of these paired markers could be used to calculate interfragmentary movement. A total of five pairs of markers were placed on the bone, three on the anterior surface and two on the posterior surface (Fig. 4). Movement of the markers was recorded onto DVtape using a Panasonic AG-SVX100 digital camcorder (Panasonic UK Ltd, Bracknell, UK). Each marker was filmed for a total duration of 20 seconds, with a load of 800N applied at a rate of 80N·sec⁻¹ over the middle 10 seconds. Before each load application, a calibration object was placed on the plane of motion which was 10mm x 10mm in size. The camcorder was focused on the markers and then zoomed in until the calibration object represented 480 pixels by 480 pixels. As a result, 1 pixel equated to 20µm in distance. The footage was transferred to a personal computer and captured using Adobe Premier Pro 1.5 (Adobe Systems Inc, CA, USA). A handwritten batch process was used to convert the movie file into a series of single frame bitmap images. The conversion was restricted to 10hz that produced a total of 100 frames per marker. By way of an in-house computer program (Digitester, University of Teesside, UK) it is possible to select each marker sequentially, throughout the full range of frames, with the pixel location in two dimensions, calculated relative to the global coordinate system defined by the calibration object, between the paired markers.

RESULTS

FINITE ELEMENT MODEL

Inferior displacement of the head of the femoral model under this 0.8kN load was 0.635mm (Fig. 5). The five screw configurations used to stabilise the fracture were analysed by comparing the peak interfragmentary movement occurring at the fracture surface of a 50° osteotomy, as this will indicate the most unstable fracture. The movement calculated is a three-dimensional representation of the fracture gap increase. The overall interfragmentary movements occurring at the fracture site were consistently less than 1mm. The lowest interfragmentary movement was found for the triangular screw configuration with a superior apical screw (0.0446mm) (Fig. 6). For the triangular screw configurations, there was little measured difference in interfragmentary movement. The range of
peak interfragmentary movement was 0.0446mm to 0.0482mm. The linear screw configuration was the least stable fracture interface with interfragmenary values of above 0.06mm.

**MECHANICAL TEST RIG**

The average axial displacement displayed by the composite femur is 0.66mm (Fig. 5), with a range of 0.6mm to 0.75mm. The force-displacement curve indicates that the composite femur displays a linear behaviour up to 0.8kN. The increase in fracture gap, determined by the maximum relative motion, fell within those defined as safe (< 5mm) [16]. Interfragmentary movement was greatest in the superior region of the femoral neck compared to the inferior region on both the anterior and posterior sides. The largest increase occurred at the superior marker on the posterior of the femoral neck (RM = 4.89mm) (Fig. 6), with the smallest motion occurring marker on the anterior inferior region (RM = 0.43mm). The tightening of the screws resulted in a reduction in the relative motion between the fracture fragments for all the marker locations. Tightening the screws by one and two revolutions had a similar effect and decreased the relative motion by an average of 56%, while a half revolution decreased the relative motion by 16%.

![Figure 5: Inferior axial displacement of the femoral under 0.8kN load.](image5)

![Figure 6: Interfragmentary movement recorded at the fracture interface dependent on techniques.](image6)
DISCUSSION

The aim of this study was to compare two in vitro techniques to measure interfragmentary movement in fixated femoral neck fractures. The techniques were a finite element model and a motion capture analysis using a mechanical test rig.

A uni-axial load was used to determine the mechanical behaviour of each model, with similar inferior displacement of 0.635mm and 0.66mm recorded (Fig. 5). This indicates a general agreement between the two types of techniques in measuring movement in femur models. The displacements were also found to be similar to the in vitro behaviour of composite femurs reported in the literature [20]. Therefore, the finite element model and mechanical test rig were considered valid. However there was an obvious difference in interfragmentary movement recorded by each technique (Fig. 6). The finite element model produced lower movement at the fractured surface compared to the mechanical model. This can be attributed to the direction of the mechanical forces which create the natural compressive effect across the fracture and increase the stability. While the movement recorded in the mechanical model is greater than that of the FE model, it is similar to other mechanical studies [16-18] and highlights the effectiveness of the motion capture technique as a useful tool for measuring the stability of fixated fractures.

A finite element model of the proximal femur was developed based on accurate geometry derived from the Visible Human Project. Region-specific Young’s modulus was applied to the model based on the gray-value of the CT images (Fig.1) [28]. This allowed a representation of the trabecular structure of the cancellous bone. This design feature is more accurate than the use of general material properties to define cortical and cancellous bone often found in previous studies. With any finite element analysis, simplifications are often made. The most obvious of these was the modeling of the screws as rigid links. Such an approach assumes infinite strength and stiffness of screws when compared to bone. Given that the titanium and stainless steel screws commonly used in surgery are over 100 times stiffer than bone and that the screws rarely fail in vivo, instead the bone fails [4, 6], and such an assumption seems justified. Furthermore, the FE model allowed the manipulation of parameters to analyse different screw configurations without the need for multiple test specimens. This is an advantage over the mechanical model that requires multiple composite femurs to test each screw configuration or test setting.

The experimental test procedure was used to measure the effect of screw tightening on the stability of an undisplaced osteotomy of the femoral neck, fixated using a triangular configuration of cannulated screws with a superior apical screw [5]. The experimental procedure maintained the complex spiral geometry of the screw thread and its interaction with the surrounding bone, which was not possible using a finite element model [4, 6, 32-33]. However, it was not possible to include internal forces to the mechanical model, due to the simplified nature of the compression test. Measurement of the interfragmentary movement was achieved by the recording of markers on the external surface (Fig. 4). The limitation of this is that it is not possible to determine the interfragmentary movement on the internal surface of the fracture interface, where the movement may have been greater.

In summary, both techniques were able to produce valid measurements of movements in a model of the femoral neck and the maximum amount of interfragmentary movement at the fracture interface similar to those reported in the literature, while the choice of which technique to use must be based on the advantages and limitation with consideration to the factors to be analysed.

REFERENCES


Evaluating Residual Life using Failure History of a Machine Tool

B. V. Sudheer Kumar*1, C. S. P. Rao2, A. Neelakanteswara Rao3

1Research Scholar 2Professor 3Assistant Professor
Department of Mechanical Engineering
National Institute of Technology
Warangal, Andhra Pradesh, 506004, India

ABSTRACT

The ever-increasing global competition in the worldwide markets is forcing the manufacturing industries to optimize the maintenance decisions. The most organized and the cost efficient strategy is preventive maintenance in many cases, but the optimal instant to perform the maintenance has to be identified. The optimal instant can be obtained by using the residual life of the system, which can be calculated through statistical analysis of failure data. If the approximate time, when failures occur were known, maintenance planning would be much simpler. Even though few techniques are in existence, still it is a challenge to determine the residual life of complex systems like machine tools having several failure modes that are dependent on age and working conditions.

In this paper, the authors presented a model for residual life estimation that uses the machine tool failure history. The failure history is used to study the failure trend by Laplace test statistic. Goodness of fit test (Cramer-von Mises test) was employed to assess the deviation of the collected data from the theoretical Non Homogeneous Poisson Process. The parameters of the Non Homogeneous Poisson Process in which failure intensity function following Weibull hazard rate function were estimated by finding the maximum likelihood estimates. Consequently, the failure intensity function can be developed for further calculations of residual life. Software was developed to find the failure intensity function from the failure data analysis and the mean failure time of the distribution of the machine tools. The software is also used to indicate future failures of the system with error bounds. The developed model was tested thoroughly on 60 machines in a shop floor of a defense industry.

Nomenclature & Acronyms

NHPP – Non-Homogeneous Poisson Process
t1 – The system age at the 1st failure
tn – the system age at the nth failure
n – the total number of failures recorded
MTTFi – instantaneous mean time to failure
u(t) – Failure intensity function
xi – the time between ith and (i-1)th failure
T – the total observed time period
Ul – Laplace test statistic
R(t) – System reliability
\( \beta \) - Shape parameter
\( \lambda \) - Scale parameter
CM – Cramer-von Mises test statistic
E[N(t)] – expected number of cumulative failures up to time ‘t’

1. INTRODUCTION

To remain competitive within the increasing global competition, the manufacturing organizations targeted maintenance budgets to reduce the total manufacturing expenditure because the direct and indirect activities associated with manufacturing processes are optimized. Typically reduction in maintenance budget may result in lesser equipment availability and reliability. The reduction in budget that increases the risk of availability of equipment will be a business strategy for many organizations. With the onset of substantial equipment availability requirements and rising costs, some organizations realized that this strategy would not produce a cost effective solution. In order to meet the competitive challenges in the manufacturing industry, a definite shift from historical maintenance practices to an advanced maintenance strategy is required. This desired program must be aimed to reduce maintenance costs while increasing the availability and reliability of our aging equipment.
A model was proposed for residual life estimation that uses the machine tool failure history. The failure history of the machine tool was used to study the failure trend whether increasing, decreasing or constant by Laplace test static [1]. A Goodness of fit test (Cramer-von Mises test) was employed to assess the deviation of the collected data from the theoretical Non-Homogeneous Poisson Process (NHPP) [2]. The parameters of the NHPP with Weibull failure intensity function were estimated by finding the maximum likelihood estimates [2]. Consequently, the failure intensity function and instantaneous Mean Time To Failure (MTTF) can be developed for further calculations of residual life. Software was developed to find the failure intensity function from the failure data analysis and the mean failure time of the distribution of the machine tools. The software is also used to indicate future failures of the system with error bounds.

2. METHODOLOGY

To start our analysis we need comprehensive failure data of the machine tool. Therefore, the first step is to collect the entire historical failure data of the machine tool up to present. After collecting the failure data, the failures were divided into two groups i.e., functional and non-functional failures. Functional failures are defined as the failures, which directly affect the performance of the machine tool. Non-functional failures do not affect the performance of the machine tool but only cause stoppage of the machine tool e.g. stoppage of motor due to contact/relay problem. These failures mostly affect the downtime of a machine tool. If the maintenance procedures are good then most of the non-functional failures can be avoided. In analyzing the machine tools, we are assuming a reasonably good maintenance. Hence consider the functional failures only for the analysis. From the various available models, we considered minimal repair model for the analysis [2]. When a system fails, if it is repaired in such a way that its age and performance remains the same as it was just prior to the failure, then the system is said to be minimally repaired system. Minimal repair is the situation where the repair of a failed system is just enough to get the system operational again [2]. We assume a complex system with many failure modes and the system reliability after the repair is same as it was just before the failure. As in most of the cases, the failure pattern of the system is assumed to be a non-homogeneous Poisson process.

If the failures of a repairable system follow a non-homogeneous Poisson process (NHPP), then the number of failures and the types of repair actions taken during a period [0,t] do not affect the probability of failure during (t, t + Δt). In particular, suppose that the system fails at time t and is subsequently, repaired and put back into service. (Repair time is assumed to be negligible). According to this model, the probability of a system failure during (t, t + Δt) is u(t)Δt, and this value would be the same even if the system had not failed at time t. In practice, however, if the repair is perfect, then one would expect a decrease in failure probability after repair from its value at the time of failure. If the system is complex, consisting of many components, then the replacement of a single component may not decrease this probability significantly [3].

2.1 Laplace test

The Laplace test, also known as the centroid test, is a measure that compares the centroid of observed arrival times with the mid point of the period of observation. This measure approximates the standardized normal random variable. The main objective of Laplace test is to determine whether the pattern of failures is significantly changing with time. This can be conducted by testing a null hypothesis that the system failure pattern follows no trend. If this hypothesis can be rejected at some appropriate significance level, then it can be concluded that some level of increasing or decreasing trend of failures is occurring. The test statistic for trend is given by

\[
U_L = \frac{\sum_{i=1}^{n} X_i - T}{\sqrt{T \frac{1}{12n}}}
\]

\[1\]

Laplace test compares the centroid of the observed failure values with the mid point of the period of observation. If -1.96<\(U_L<1.96\) there is no trend, i.e. the process is stationary. If \(U_L<-1.96\) the trend is decreasing, i.e. the inter-failure values are tending to become larger. Conversely, when \(U_L>1.96\) the trend is increasing, i.e. inter-arrival values are tending to become progressively smaller [1]. If the period of observation ends at an event, use \((n-1)\) instead of \(n\) and exclude the time to the last event from the summation.
2.2 Goodness of fit test

In analyzing statistical data we need to confirm how well the data fits in the assumed distribution. The goodness of fit can be tested statistically, to provide a level of significance that the null hypothesis is rejected. Goodness-of-fit test is a statistical test in which the validity of one hypothesis is tested without specification of an alternative hypothesis.

The final step in the selection of a theoretical distribution is to perform a statistical test for goodness of fit [2]. Such a test that compares a null hypothesis (H₀) with an alternative hypothesis (H₁) having the following form:

H₀: The failure times came from the specified distribution.
H₁: The failure times did not come from the specified distribution.

The test consists of computing a statistic based on the sample of failure times. This statistic is then compared with a critical value obtained from a table of such values. Generally, if the statistic is less than the critical value, the null hypothesis (H₀) is accepted; otherwise, the alternative hypothesis (H₁) is accepted [2].

The critical value depends on the level of significance of the test and the sample size. The level of significance is the probability of erroneously rejecting the null hypothesis in favor of alternative hypothesis. There are four possible cases that may occur because of the randomness inherent in the sampling process, the test statistic has a probability of exceeding the critical value even though H₀ is true. This results in a Type 1 error having a probability of occurring equal to the level of significance (α). It is also possible for the test statistic to be less than the critical value even though H₁ is true. This results in a Type 2 error. It occurs with a probability that is usually controlled indirectly by the specification of the level of significance and the sample size.

There are two types of goodness-of-fit tests i.e., general tests and specific tests. A general test is applicable to fitting more than one theoretical distribution, and a specific test is tailored to a single distribution. When available, specific tests will be more powerful (having a higher probability of correctly rejecting a distribution) than general tests [4].

2.2.1 Cramer-von Mises test

To determine whether the non-homogeneous Poisson process is a more appropriate model than the constant failure rate model (homogeneous Poisson process), a trend test on the failure times is performed. For the intensity function \( u(t) = \lambda t^{\beta-1} \), the hypothesis tested is:

H₀: The intensity function is constant (\( \beta=1 \)).
H₁: The intensity function is not constant (\( \beta \neq 1 \)).

If the intensity function is not constant, then either an increasing trend (deteriorating), \( \beta > 1 \), or decreasing trend (growth), \( \beta < 1 \), is present and the non homogeneous Poisson process should be considered.

If a trend is present, the power-law intensity function may be suitable model. To perform the goodness-of-fit test for this model, the following hypotheses are stated [5].

H₀: The failure data follows the Non homogeneous Poisson process.
H₁: The failure data does not follow Non homogeneous Poisson process.

The Cramer-von Mises goodness-of-fit test statistic is then computed from

\[
C_M = \frac{1}{12M} + \sum_{i=1}^{M} \left[ \left( \frac{t_i}{t_n} \right)^\beta - \left( \frac{2i-1}{2M} \right) \right]^2
\]

where \( \bar{\beta} = \hat{\beta} \times [(n-2)/n] \) (2)

The calculated C_M value is compared with a critical value obtained from a table. If this C_M value is less than the corresponding critical value at a significance level we can accept the null hypothesis that the failure times of the system follows a Non homogeneous Poisson process with intensity function \( u(t) = \lambda t^{\beta-1} \).

2.3 Parameter Estimation to calculate Failure Intensity Function

The failure data collected for the machine tools are mostly failure-truncated data. Therefore only the formulae applicable for failure-truncated data are considered here. To calculate the failure intensity function \( u(t) = \lambda t^{\beta-1} \) we have to calculate the parameters \( \hat{\lambda} \) and \( \hat{\beta} \) from the following equations [5].
\[ \hat{\beta} = \frac{n}{(n-1) \ln t_n - \sum_{i=1}^{n-1} \ln t_i} \]  

(4)

\[ \hat{\lambda} = \frac{n}{T^{\hat{\beta}}} \]  

(5)

By substituting the values of the parameters \( \hat{\lambda} \) and \( \hat{\beta} \) in the equation the failure intensity function can be calculated as \( u(t) = \hat{\lambda} \beta t^{\beta-1} \)  

(6)

2.4 Calculation of instantaneous mean time to failure (MTTF) with lower and upper bounds

Instantaneous mean time to failure (MTTF) can be calculated from the formula.

\[ \text{MTTF} = \frac{1}{u(t_n)} \]  

(7)

Two sided confidence intervals for the MTTF can be obtained from the following equations.

- Lower bound = LL (MTTF) \( (8) \)
- Upper bound = UL (MTTF) \( (9) \)

The values of LL and UL can be obtained from standard statistical tables [5].

2.5 Expected number of failures

The cumulative expected number of failures from 0 to \( t_0 \) is given by [6].

\[ E[N(t_0)] = \int_0^{t_0} u(t) \, dt \]  

(10)

2.6 Calculation of Residual Life (RL)

If the load and the working conditions of the system are uniform then the residual life can be calculated from the following formula [7].

\[ \text{Residual Life} = \text{MTTF} - \text{System age from the recent failure} \]  

(11)

2.7 System reliability

The probability \( R(t) \) that a system of age \( t \) will successfully complete a mission of fixed duration \( d > 0 \) is called "System reliability." If the system is repairable and mission aborting failures follow a non homogeneous Poisson process with Weibull intensity function, then \( R(t) = \text{Probability of system of age } t \text{ will not fail in } (t, t+d) \) [3].

\[ R(t) = \exp\left[-(\hat{\lambda} (t + d)^{\hat{\beta}} - \hat{\lambda} (t)^{\hat{\beta}})\right] \]  

(12)

3. CASE STUDY OF A ROBOFIL- 310 CNC CHAR MILLER EDM WIRECUT MACHINE

Date of Installation of the Machine Tool is 2nd September 1996. For the Machine tool no theoretical normal lifetime is available. Data of the above machine is failure-truncated type (type II). Therefore for all further calculations \( T = t_n \)

3.1 Laplace trend test

From the collected data of the machine the following values are calculated.

- \( n = 50 \) (Total no. of failures)
- \( T = t_n = 2858.571 \) days
- \( U = 2.8635 \) [The value was calculated by using equation (1)]

**Conclusion:** The value of \( U \) is found to be more than 1.96. The trend is increasing, i.e. inter-arrival values are tending to become progressively smaller which indicates that the system is deteriorating.
3.2 Cramer-von Mises test

\[ \beta = 1.5727 \quad \text{[The value was calculated by using equation (4)]} \]
\[ \beta = 1.5098 \quad \text{[The value was calculated by using equation (3)]} \]
\[ M = n-1 = 49 \]
\[ C_M = 0.0588 \quad \text{[The value was calculated by using equation (2)]} \]

The critical value at 0.2 significance level is 0.129. (The value was taken from standard statistical tables.)

Conclusion: The test static \( C_M \) is found to be less than the critical value. Hence hypotheses \( H_0 \) is accepted i.e., the failure data follows the Non homogeneous Poisson process with intensity function \( u(t) = \lambda \beta t^{\beta - 1} \)

3.3 Parameter Estimation to calculate Failure Intensity Function

\[ \hat{\beta} = 1.5727 \quad \text{[The value was calculated by using equation (4)]} \]
\[ \hat{\lambda} = 0.0001834 \quad \text{[The value was calculated by using equation (5)]} \]
\[ u(t) = 0.0002884t^{0.5727} \quad \text{[The value was calculated by using equation (6)]} \]
\[ u(T) = 0.0275 \]

<table>
<thead>
<tr>
<th>t</th>
<th>u(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.0101</td>
</tr>
<tr>
<td>1000</td>
<td>0.0151</td>
</tr>
<tr>
<td>1500</td>
<td>0.0191</td>
</tr>
<tr>
<td>2000</td>
<td>0.0224</td>
</tr>
<tr>
<td>2500</td>
<td>0.0252</td>
</tr>
<tr>
<td>2858.6</td>
<td>0.0275</td>
</tr>
</tbody>
</table>

3.4 Calculation of instantaneous mean time to failure (MTTF) with lower and Upper bounds

Instantaneous mean time to failure (MTTF) can be calculated from the formula

\[ MTTF_i = 36.3636 \text{ days} \quad \text{[The value was calculated by using equation (7)]} \]

Two sided confidence intervals for the MTTF can be obtained from the following equations.

Lower bound = 29.298 days \quad [\text{The value was calculated by using equation (8)}]
Upper bound = 49.164 days \quad [\text{The value was calculated by using equation (9)}]

(The values of LL and UL can be obtained from the standard reliability tables.)

Conclusion: The instantaneous MTTF value is 36.3636 days. The lower bound is 29.298 days and the upper bound is 49.164 days with a confidence level of 0.80.

3.5 Expected number of failures

\[ E[N(3000)] = \int_0^{3000} 0.0002884t^{0.5727} \, dt = 53.93 \approx 54. \quad \text{[The value was calculated by using equation (10)]} \]

Conclusion: In the next 3000 – 2858.6 = 141.4 \approx 141 days the expected number of failures are 54 - 50 = 4
3.6 Calculation of Residual Life (RL)

By assuming uniform load and the working conditions, the residual life of the machine tool can be calculated from the following formulae.

\[ \text{Residual Life} = \text{MTTF}_i - \text{System age from the recent failure} \]

If the system age is assumed as 10 days from the recent failure

\[ \text{Residual life} = 36.3636 - 10.0000 = 26.3636 \text{ days} \quad \text{[The value was calculated by using equation (11)]} \]

3.7 System reliability

The reliability of the machine was calculated for the next 60 working days (d) after the observed time period (T). A graph was plotted between reliability and time. The reliability values were obtained from equation (12) for various values of time t.

<table>
<thead>
<tr>
<th>t</th>
<th>R(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.3722</td>
</tr>
<tr>
<td>1000</td>
<td>0.2355</td>
</tr>
<tr>
<td>1500</td>
<td>0.1630</td>
</tr>
<tr>
<td>2000</td>
<td>0.1185</td>
</tr>
<tr>
<td>2450</td>
<td>0.0914</td>
</tr>
<tr>
<td>2858.6</td>
<td>0.0735</td>
</tr>
</tbody>
</table>

3.8 Results & Analysis

Generally, any statistical model can make good short-term predictions only. For long-term predictions, no model will be trustable because the system may change over time. The best way to make better predictions is to add the new failures using the new observations dynamically.

The data considered for Robofil CNC machine tool contains fifty entries. To evaluate the method developed, 80% of the data was considered to estimate the future failures similar to the artificial neural networks. The estimated failures were compared with the remaining 20% of the actual failures to find the efficiency of the developed method. Among the fifty entries, 40 entries were considered and the 41st failure was estimated. 42nd failure was estimated by considering 41 failures. 43rd failure estimate by 42 failures, 44th failure estimate by 43 failures and so on.

41st failure estimate = 43.24 days  
42nd failure estimate = 42.38 days  
43rd failure estimate = 41.50 days  
44th failure estimate = 40.75 days  
45th failure estimate = 39.21 days  
46th failure estimate = 37.61 days  
47th failure estimate = 36.85 days  
48th failure estimate = 36.63 days  
49th failure estimate = 36.67 days  
50th failure estimate = 36.07 days

(The values of LL and UL are taken from standard statistical tables)

At 0.90 confidence level LL= 0.7303 & UL = 1.538 (for n=41 to 45)
At 0.90 confidence level LL= 0.7513 & UL = 1.46 (for n=46 to 50)
At 0.95 confidence level LL= 0.6832 & UL = 1.66 (for n=41 to 45)
At 0.95 confidence level LL= 0.7076 & UL = 1.562 (for n=46 to 50)
At 0.98 confidence level LL= 0.6328 & UL = 1.816 (for n=41 to 45)
At 0.98 confidence level LL= 0.6605 & UL = 1.692 (for n=46 to 50)
Table 1: lower and upper bounds for 0.98 confidence level

<table>
<thead>
<tr>
<th>Failure No</th>
<th>Lower bound (in days)</th>
<th>MTTF (in days)</th>
<th>Upper bound (in days)</th>
<th>Actual failure time (in days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>27.36</td>
<td>43.24</td>
<td>78.52</td>
<td>30.71</td>
</tr>
<tr>
<td>42</td>
<td>26.82</td>
<td>42.38</td>
<td>76.96</td>
<td>28.57</td>
</tr>
<tr>
<td>43</td>
<td>26.26</td>
<td>41.50</td>
<td>75.36</td>
<td>30.71</td>
</tr>
<tr>
<td>44</td>
<td>25.79</td>
<td>40.75</td>
<td>74.01</td>
<td>08.57</td>
</tr>
<tr>
<td>45</td>
<td>24.82</td>
<td>39.21</td>
<td>71.21</td>
<td>07.14</td>
</tr>
<tr>
<td>46</td>
<td>24.84</td>
<td>37.61</td>
<td>63.64</td>
<td>20.00</td>
</tr>
<tr>
<td>47</td>
<td>24.34</td>
<td>36.85</td>
<td>62.35</td>
<td>39.29</td>
</tr>
<tr>
<td>48</td>
<td>24.19</td>
<td>36.63</td>
<td>61.98</td>
<td>46.43</td>
</tr>
<tr>
<td>49</td>
<td>24.22</td>
<td>36.67</td>
<td>62.05</td>
<td>27.14</td>
</tr>
<tr>
<td>50</td>
<td>23.82</td>
<td>36.07</td>
<td>61.02</td>
<td>52.86</td>
</tr>
</tbody>
</table>

For the above considered example
When the confidence level is 0.90 three (3) actual failure times i.e., 30% are matched with the estimated values.
When the confidence level is 0.95 six (6) actual failure times i.e., 60% are matched with the estimated values.
When the confidence level is 0.98 seven (7) actual failure times i.e., 70% are matched with the estimated values.
Hence the methodology developed was found to be reasonably good.

4. CONCLUSIONS

- With the help of prior information regarding the failures, maintenance planning will be more accurate, which increases the availability of the machine tool.
- By analyzing the reasons for the failure with highest frequency the critical subsystem / component of the machine tool can be identified, which can be useful in improvising the design.
- By estimating the number of failures in a given period the inventory can be procured as per the demand, which can reduce the inventory cost.
- Reduction in downtime of machine tool results in increased productivity and reduction in production cost.
- From a set of exactly similar machine tools to select the most reliable machine tool to manufacture critical components, the system reliability plays a major role in the selection.

REFERENCES

Maintenance Management Strategy Development: An Explorative Study in a Large Oil Company

Jalal Ashayeri*, Daan F.J. Schraven†

Dept. of Econometrics & Operations Research
Tilburg University
P.O. Box 90153, 5000 LE, Tilburg
The Netherlands

ABSTRACT

As momentum continues worldwide to hold down business costs, large manufacturing companies are looking at every option available to reduce expenses. They outsource engineering and management capabilities like maintenance operations, including purchasing management of spare parts. In an effort to track the strategic impact of such practices at different consideration levels like: supply chain, resources, and operations, this paper summarizes an exploratory research of a real life case-study at a manufacturing plant of a large oil company in the Netherlands. The paper presents a systematic approach of analyzing the maintenance requirements in order to assist management to develop proper proactive maintenance strategies for the manufacturing assets. This approach highlights critical parameters for evaluation and tracking that are crucial for adapting the right maintenance strategy.

1. INTRODUCTION

With the advent of global recession, many large companies are faced with what seems to be an impossible situation – maintaining or improving manufacturing reliability while working with reduced staffing and declining maintenance budgets. Maintenance however has an important role in the overall operations of a business success. In fact, the strategic relevance of the maintenance function is increasing [1]. This covers both intra and inter-organizational levels. Inside an organization, maintenance is capable of significantly influencing the income statement [2], e.g. performance and cost control. In a supply chain, the strategic objectives (like being agile, lean, or flexible) impact on the system of the maintenance management activities. Despite these facts, a growing number of companies are outsourcing maintenance operations, management control, including purchasing of maintenance related materials. Therefore, given the supply chain strategy manufacturing plants require to formulate a maintenance strategy and to implement it at different levels.

The strategy formulation process for maintenance management is a well-known subject in the literature. Among others, frameworks have been developed to add value [3], increase competitiveness [4] and establish profitability within maintenance operations [5]. Moreover, all sorts of innovative approaches to maintenance problems exist nowadays [6]. Total productive maintenance, preventive maintenance and reliability centered maintenance are popular examples in the long list. These concepts all aim at improving the effectiveness of production equipment to eventually advance productivity levels [6].

Due to the situation specific nature of maintenance such as unique equipment properties of a manufacturing plant; particular process requirements; existing company-wide policies and strategies; management priorities, numerous strategy development process frameworks and concepts can be found in the literature. In many real-life situations however, management is not aware of all concepts and tools available and all dimensions involved in a case. Therefore, the strategic focus would be based on the requirements and goals that plant management prioritizes and understands. Moreover, this prioritization is often subject to a spending budget.

* Corresponding author, J.Ashayeri@uvt.nl
† Currently at Dept. of Construction Management & Engineering, University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands
This paper summarizes an exploratory research of a real life case-study at a manufacturing plant of a large oil company in the Netherlands. The paper presents a controlled priority break-down for the maintenance spending budget for managing multiple resources simultaneously in a reliability based maintenance environment. This entails developing a systematic way of analyzing the requirements of the manufacturing plant and management. Section 2 of this paper identifies the plant requirements. Section 3 discusses the adapted study approach. The approach is based on a combination of Economic Optimum and Reliability Assurance to define the criticality of assets. Techniques such as Failure Mode and Effect Criticality Analysis (FMECA), Risk Priority Numbers (RPN), Fuzzy Set Analysis, HAZOP, Multi-criterion Method, and Analytical Hierarchy Process (AHP) are used for this purpose. Section 4 provides the key findings of the analysis. Section 5 concludes the paper.

2. COMPANY BACKGROUND & THE CHALLENGE

The plant produces multiple lube oils for applications in shipping, aviation, industry and road transport. The production process comprises two major processes. First, the blending process is performed to mix base oils with additives to create products according to a recipe. This contains an inline blending facility, an automated process where 10 loops are connected that blend end-products through circulation. Another facility is the batch blending facility, a semi-automated process where additives are poured into pits for blending in a series of connected tanks. Second, the filling process is performed to pack the products in different sized containers: drums, kegs or pails for transportation. These packed products are then stored in an adjacent warehouse.

At the plant some maintenance challenges are observed. First, some equipment, e.g. inline blending facility is rather aged while it is an important part of the production process. This increases wear-out risk. Second, sometimes, spare parts are not readily available during repairs. This causes delay for production. Third, the entire plant is expanded during building phases over a period of fifty years. This has led to a high diversity of equipment characteristics. Outsourcing maintenance in this environment is in effect handing over control of these potential sources of risk to an external party. However, it may induce improvements. Therefore, the company has subcontracted major parts of maintenance activities to a 3rd party maintenance service provider (3PMSP). The results however were not satisfactory and initiated this study.

Like other process industries, the plant performance here is highly dependent on its physical assets and should view maintenance as a necessity [7]. However, not all physical assets require the same attention where dependability is concerned. Therefore, the criticality of the physical assets with respect to production contribution is a decisive factor for the type of strategy. Plant management was concerned about the results of outsourcing and the strategic direction of the maintenance activities. Therefore, the study aimed at developing a systemic approach to assist management determining the strategic directions. It took into account that the approach should focus on improving reliability assurance of the equipment and optimizing the budget spent for all maintenance operations and its related activities. Interviews were conducted to identify the challenges and opportunities, which are summarized in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Opportunities</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability management</td>
<td>(1) Reliability data is partially available and partially retrievable with policy adjustments at the lube plant.</td>
<td>(1) Maintenance operational performance is not effective by providing high reliability assurance in preventing and predicting breakdowns at the lube plant.</td>
</tr>
<tr>
<td>Maintenance Operations</td>
<td>(2) The preventive maintenance schedule does not include performance improvements; it has basic requirements enforced by the government and corporation.</td>
<td>(2) The lube plant is ignorant of applying reliability assurance for a decision-making system on maintenance-related purchasing.</td>
</tr>
<tr>
<td>Maintenance related Purchasing</td>
<td>(3) There are advantages to be explored by a renewed situation in contracting the maintenance service at the lube plant.</td>
<td>(3) The lube plant is occupied with corrective maintenance for the majority of the time.</td>
</tr>
<tr>
<td></td>
<td>(4) Uncertainty exists in the dynamics of purchasing MRO items for different types of equipment at the lube plant.</td>
<td>(5) Uncertainty exists in the way to design and implement appropriate capital replacement of existing equipment.</td>
</tr>
</tbody>
</table>
2.1 RELIABILITY MANAGEMENT

Managing reliability and maintaining equipment coherently is important for the competitiveness of an organization [4]. An implementation of a reliability scheme can be performed by means of a reliability program, e.g. to steer maintenance operations [8]. Equipment maintenance in this way involves maintenance operations and purchasing strategies [9]. Such a scheme does not exist at the plant. The plant does have an asset management system to enable this opportunity. Clear directional vision, life cycle analysis, maintenance strategy, and risk management are required for such system to succeed. However, this information is not fully disclosed by 3PMS to the plant for strategic reasons. Moreover, the existing monitoring and data collection system is not well suited for reliability management since this data is partially available and retrievable. Therefore, the asset management system is an opportunity, but somehow must be made more transparent and extensively useable to plant management by the 3PMS.

Reliability management opportunity has two challenges. First, the maintenance operations at the plant do not have a strategic intent. The operational performance is not effective by providing high reliability assurance in preventing and predicting breakdowns at the plant. It has no testing infrastructure to control breakdowns. Second, plant management does not have an understanding of the reliability concept. Therefore, they do not have the knowledge of how to include it in their decision making. Concerning the maintenance-related purchasing, the reliability of equipment cannot be assured.

2.2 MAINTENANCE OPERATIONS

In terms of maintenance operations the plant employees are occupied with corrective maintenance and firefighting for the majority of their time. There is a schedule with preventive maintenance activities in place for safety measures and basic requirements enforced by policy, e.g. the government. Thus operationally a comprehensive scheduling system is required to increase reliability [6].

2.3 MAINTENANCE-RELATED PURCHASING

The maintenance operations have been mainly outsourced for a few years prior to this study. The outsourcing limits the strategic control of plant management. In literature it is highly debatable whether outsourcing maintenance is a good step [10; 11]. The advantages for a renewed contracting situation in the maintenance service can be explored.

The lube plant is exploring a way to manage their inventory of spare parts effectively. There are numerous possibilities that can contribute to a good solution: spare parts inventory at an optimal quantity, keeping inventory at the supplier and supply base optimization. However, every type of equipment has its’ own properties. Therefore, uncertainty exists in the dynamics of purchasing spare items for different types of equipment.

The lube plant has been built in the 1950s. Since then, several times the plant has been expanded and partially rebuilt. As a result, the equipments have divergent operating ages. However, no historical data is kept within the maintenance system. The knowledgeable employees have left the company and those remained have no recollection of specific events in the past. In the mean time, the aging equipment can wear out any time. However, the production process depends highly on their performance. Therefore, uncertainty exists in the way to design and implement appropriate capital replacement of existing machines.

In essence the plant lacks a clear strategy definition and requires a systematic approach to evaluate the consequence of strategic maintenance focus.

3. THE APPROACH

Supply chain goals are first translated into manufacturing targets and in turn these are translated into maintenance objectives which then form the basis of maintenance strategy formulation. Subcontracting part or whole maintenance operations is the result of this process. Conventionally outsourcing strategy suggests that one should outsource the “non-core” business activities. When it comes to maintenance strategy, the difficulty with this approach, however, is that there is no guidance for deciding which maintenance activities are “non-core”. Ultimately, the discussion about what is “core” and what is “non-core” depends on the maintenance objectives. The approach presented in Figure 1 is meant to answer these interacting questions.
A few critical parameters for evaluation and tracking of the impacts of management decisions are considered in this model. First, criticality measurements are used in determining the key machines for maintenance attention [12; 13]. Second, the arguments of an economic optimum and reliability assurance regulate the way a prioritization of the equipment is managed. Third, after a criticality analysis the findings are often presented to management [12] and are checked whether they are aligned and in line with overall supply chain strategy. They either confirm or partially adjust the criticality results. Fourth, for each individual piece of equipment, strategies of each discipline are often assigned [12], again in line with the supply chain strategy adopted for the plant. The approach therefore yields different decision nodes: (a) the criticality rating decision, which technique is the most optimal and informative to use, subject to equipment properties. This study treats five techniques that are explained below. (b) The management prioritization decision, which set of equipment (activities), requires highest level of attention, given the criticality results, certain judgmental criteria, and the supply chain strategic objectives. The judgment criteria are explained below. (c) Finally the maintenance strategic focus decision is made which separates basic and strategic maintenance requirements.

3.1 CRITICALITY RATING DECISION

The choice of criticality measures relates to the equipment scope and scales and whether the use of measure is economical to review and rate criticality. For this study the economical optimum is defined as the wide-scale and universal application of the method. We define the reviewing process as the degree of detail one retrieves from single equipment. Five methods are considered and analyzed in this context. The study considered two types of equipment: a population of 20 randomly selected pumps and a single 20 liter filling line. In total 13 variables are used for quantitative analysis of the pumps. One independent study on the 20 liter filling line was used for comparison reason. These methods are briefly defined along with the literature.

3.1.1 The Failure mode and effect criticality analysis

This is a method that rates any machine sample according to a stepwise procedure. Our analysis is based on the work of Bertolini and Bevilacqua [14] and Moss and Woodhouse [15]. Two types of FMECA processes have been considered for this study: Risk Priority Numbers (RPN) and Fuzzy Set studies. These are briefly distinguished below. In short, the start procedure is identical for both techniques. For each separate element of an arbitrary number
Maintenance Management Strategy Development: An Explorative Study in a Large Oil Company

of machines a description must be given of 1) function; 2) failure cause and mode; 3) the effect per failure mode; 4) side remarks.

RPN: a quantitative assessment is done through RPN. We used this method based on the work of Kmenta and Ishii [16] for the 20 pumps. The method uses three recognized variables that are rated on an interval scale of 1 to 10: 1) severity; 2) occurrence; 3) and detectability. The formula is given below. This distinguishes the criticality of multiple pumps in an index of 1 to 1000. We surveyed production and maintenance operators that worked frequently with these devices. They were asked to rate statements about the variables on an interval scale of 1 to 5 for each pump. They marked 0 where no experience was the case. Blank responses were treated by substituting for the average value.

\[ Criticality = RPN = x_{severity} \cdot x_{occurrence} \cdot x_{detectability} \]

Fuzzy Set: this qualitative assessment uses expert knowledge from the beginning of the analysis to describe the failure modes and effects in more detail. For each separate component of a machine more detail is checked if required. This study was conducted, independent of this paper, by a maintenance engineer for a 20 liter filling line in the plant. We compare this with the RPN, focusing on the equipment differences.

3.1.2 Hazard and operability study (HAZOP)

This type of analysis identifies the ways process equipment can malfunction or deviate [17]. This analysis is more technical in respect. It is not based on experience with equipment, but on knowledge about the equipment. The company version of this analysis is used to test the HAZOP procedure. We verified the principles with the work of Cotaina et al. [17] to guard the theoretical definition of the HAZOP. HAZOP analyzes two dimensions: the likelihood of deviations and severity of consequences, see the formula. We surveyed two plant engineer to rate 5 variables on category scales of 1 to 4 per pump. These variables can be found Appendix A; row-numbers 4 to 8.

\[ Criticality = x_{consequence} \cdot x_{probability} \]

3.1.3 Multi-criterion method using multiple regression (MCM MR)

This method is introduced by Hijes and Cartagena [18]. The technique uses multiple variables that like RPN, reliability and cost metrics. Our MCM method for the 20 pumps is based on Hijes & Cartagena [18] and Keller & Warrack [19]. The multiple regression analysis is displayed below. We use the RPN method to define the dependent variable: criticality. We surveyed plant engineers and operators to get 10 independent variables. These variables can be found Appendix A; row-numbers 1 to 10.

\[ Criticality = \alpha + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_nx_n \]

\( \alpha \): interceptor.
\( \beta \): coefficient of variable X.
\( n \): amount of variables.

3.1.4 Multi-criterion Method with Analytical Hierarchy Process (AHP)

This is an approach that uses subjectively stated criteria in normalized quantitative weight factors for each independent variable. We simulate this MCM based on Bevilacqua and Braglia [12]. Again we identify criticality by using the RPN. The 10 independent variables are weighted for importance using a matrix and discussing the interrelations.

\[ Criticality = \omega_1x_1 + \omega_2x_2 + \ldots + \omega_nx_n \]

\( \omega \): weight factor of variable.

3.2 MANAGEMENT PRIORITIZATION DECISION

The typical management prioritization decision is oversimplified. It is static. It is typically based on naming the priority, listing the priorities in sequence without classifying properly the implementation issues. We need to evaluate how resources are used and identify risk of total or large-partial failure. For this purpose we considered, in addition to the equipment criticality, certain judgment criteria and the supply chain strategy of the business line. In terms of the judgment criteria we adopted questions from reliability and maintenance turnaround programs of Backlund [20] to be posed to the management. These questions include:

- Change management: Gradual or revolutionary implementation?
- Small scale pioneering: is a pilot study useful to gain experience beforehand?
- Focus areas: whether the area considered generates a company-wide improvement?
Value of analysis: should criticality results be an indicator or decision maker?

With respect to the supply chain strategy we analyzed the business line supply chain strategy statement of the company. Questions that we posed to plant management include:

- How can the business line supply chain strategy be translated to plant level?
- How can the maintenance budget allocation be determined by means of criticality intelligence?

The advantages of this approach for prioritizing are many, among those one could indicate, increased probability of avoiding failure and successfully meeting objectives, more flexibility of action when objectives change, and more flexibility of using resources actually available, resulting in increased efficiency.

3.3 STRATEGIC FOCUS DECISION

Numerous maintenance strategies are possible for every business [13]. Some require basic treatments and not much management focus, e.g. corrective maintenance. Others are strategic and require a lot of emphasis and investments, capital replacements. Each category of Table 1 can therefore have multiple treatments with different intensity levels. These treatments differ because of equipment characteristics in the maintenance field [12, 14] or the purchasing field [21, 22].

We use a portfolio approach from literature to frame each treatment within the categories of Table 1. A portfolio approach is characterized by its relational alternatives [22] and matrix structure [23]. This way options can be structured in a matrix for decision making. For example, one of the five portfolio approaches used considers the outsourcing maintenance decision along two dimensions. The first, Strategic-Non Strategic, considers how important the activity proposed for outsourcing is to the organization in achieving long term strategic competitive advantage. In terms of maintenance, this will clearly vary from organization to organization, depending on the industry that it competes in, and its chosen strategy for competing in that industry. The second dimension, Competitive-Non Competitive, relates to how competitively the function being considered for outsourcing is currently being performed compared to the external competitive environment. This relates primarily to the cost of the service, but could also be extended to include service elements such as response time. This provides strategic intent for every option under study [24].

4. RESULTS & DISCUSSION

The results of the study based on the approach presented are given and discussed in this section.

4.1 RESULTS OF CRITICALITY RATING DECISION

The results of the RPN FMECA analysis are summarized in Table 2. The mean score was multiplied by factor 2. The population represents the amount of average responses on all the pumps. The mean of severity is remarkably high. The standard deviation of occurrence and detectability are also high. The population of detectability is low, since a small group of maintenance operators were questioned for this variable only.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity</td>
<td>8.37</td>
<td>1.84</td>
<td>8.33</td>
</tr>
<tr>
<td>Occurrence</td>
<td>5.15</td>
<td>2.29</td>
<td>11.00</td>
</tr>
<tr>
<td>Detectability</td>
<td>4.93</td>
<td>2.28</td>
<td>2.75</td>
</tr>
</tbody>
</table>

The comparison study of, among other, the RPN and Fuzzy Set is summarized in Table 4. The RPN was conducted through a questionnaire provided to operators. The statements about the variables were broad. The quantification was conventional. We had a relatively large sample size, 20 pumps. The purpose was also to rank the pumps amongst each other. We did not break down the pumps into parts to rate them.

The Fuzzy Set study made use of expert judgment by a single engineer. The engineer wrote the report by analyzing the failure causes. Moreover, he broke down the filling line into subsystems identify the hazardous components. He used discussion, reason and visual techniques, e.g. modeling and pictures to draw conclusions about causes.

The HAZOP analysis of the internal company policies is displayed in Figure 2. The two types of risk (see section 3) are displayed: consequence risk and probability risk. A risk of 1.0 is critical. A score of 4.0 is non-critical. The
pumps are not scattered over the field, indicating a problem with the way the analysis was conducted. The category scale of the internal HAZOP was higher level focused than one single pump could substantially contribute to the risk definition. On the other hand, the pumps are technically identical.

Figure 2 – Scatter Diagram of the HAZOP

Table 3 shows the top 5 of the numerical ranking for the four rating techniques. The ranking is based on the highest scores, reversely for the HAZOP. This top 5 of most critical pumps is identical for all the methods. From the Pareto principle the results seem fairly stable for each method.

Table 3: Comparison of 4 ranking methods.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Pumppp</th>
<th>RPN Score</th>
<th>Pump</th>
<th>HAZOP</th>
<th>Pump</th>
<th>MCM/MR</th>
<th>Pumppp</th>
<th>MCM/AHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>540.00</td>
<td>7</td>
<td>11.38</td>
<td>7</td>
<td>389.19</td>
<td>7</td>
<td>5.74</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>400.00</td>
<td>8</td>
<td>11.38</td>
<td>9</td>
<td>364.07</td>
<td>12</td>
<td>5.64</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>360.00</td>
<td>9</td>
<td>11.38</td>
<td>8</td>
<td>353.92</td>
<td>13</td>
<td>5.15</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>350.00</td>
<td>6</td>
<td>12.25</td>
<td>12</td>
<td>352.82</td>
<td>9</td>
<td>5.08</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>300.00</td>
<td>11</td>
<td>12.25</td>
<td>13</td>
<td>342.67</td>
<td>8</td>
<td>4.76</td>
</tr>
</tbody>
</table>

Table 4 shows a comparison of the ranking methodology characteristics. We generated this information after conducting the analyses. The facts and resources are unique for this paper only. The purpose and properties might be useful to other applications of these methods.

We observe a low number of variables and respondents in the HAZOP analysis. The group of engineers with sufficient pumps expertise was small. With the more conventional approaches like RPN and MCM, the number of respondents is large.

The purpose of study types differ as well. The conventional approaches focus more on rating the functional judgment with statistics. The HAZOP and Fuzzy set rely more on engineers to value technical data about the machines. The fuzzy set is more in-depth focused. The HAZOP is more of a general overview with more technical emphasis.

The resources used to follow the procedures are related to the way the variables are measured. In appendix A the variables and their measurement are given. The RPN study used only responses to the questionnaires. The report of the filling line indicates the use of a field study. Furthermore the HAZOP relied on the instant knowledge of the engineers and their response. Since both the MCM studies relied on most of the variables, both are used. The Fuzzy set study has a high use of qualitative methods, since discussions and visual techniques are dominant tools.

The properties mainly pinpoint the advantages and disadvantages of each approach. The RPN study can easily be used for a large machine sample to get a broad overview by experience-based rating. In contrast, the Fuzzy set study is effective on a small machine sample for a detailed picture of failure causes from engineer expertise. The HAZOP also relies on the engineering expertise. However, this analysis is just as effective for a large machine sample and it is more in-depth than the RPN. Both the MCM studies are quite identical in properties. The amount of variables included in the formula is highly dependent on the effectiveness for large samples or repetitions. The only difference in terms of approach that is evident is the way the weight factors are determined. The AHP uses pair-wise
comparison which is less complex [12]. The multiple regression uses direct incorporation of all quantitative variables which is more statistically exhaustive [18]. In general the MCM methods are both effective to use when variable data is easily retrieved and a wide scope rating is required.

The choice of criticality rating method is depends not only on the plant requirements in general, but also on the initial data availability of equipment and the sources needed to collect and measure. For a comprehensive analysis an interactive implementation starting at the more conventional methods going into more detailed one is advisable. For a specific purpose of broad or detail picture enhancement each of the individual methods can be screened.

### 4.2 RESULTS OF MANAGEMENT PRIORITIZATION DECISION

In this section we discuss the input factors from Figure 1 for the plant management prioritization task. These include: criticality rating analysis, judgment criteria and the supply chain strategy.

The criticality rating analysis does not result in absolute decisions. The various equipment conditions rating brings to light areas requiring more attention, and therefore management should set the priorities. As was intuitively expected, reliability improvement of set of pumps is a priority in this case. We observed from the HAZOP pump study, Figure 2 that the risk levels are not significantly scattered. Moreover, the comparison of the criticality methods, Table 4, suggests a useful sequence for analyzing the assets progressively in-depth. Therefore, we propose that the criticality tools that are used company-wide, e.g. HAZOP, need to be personalized in the risk levels appropriate for the type of asset. Moreover, we propose an active feedback loop to gather further intelligence on a more qualitative scale for those assets that appear highly critical from a larger sample.

Concerning the judgment criteria there are some theoretically correct answers to the posed questions. Change management is more advisable in an evolutionary approach when reliability is an important consideration factor. Ansell et al. [25] recognizes that a gradual growth in reliability is beneficial. In fact, large scale improvement projects have often more failed than succeeded [20]. In that respect a short term incremental project that yields tangible results is advisable [26]. Such an incremental project can be a pilot study as proposed by Backlund [20]. In such a pilot study a small scale pioneering is achieved through acting as if it was an official strategy or policy implementation. In order to prepare such a pilot study, an initiation phase requires four objectives to be met [20]. First, a maintenance evaluation on three levels needs to be achieved: company-wide [27]; equipment; and performance measurements [28]. Second, the maintenance management strategy requires being formulated consistently and feasible. Third, the potentials of the pilot study must be defined [20]. This way a clear vision is created about the expected achievements. Fourth, the performance indicators are practically defined according to the
strategy and expectations. In the pump study these questions did not get a satisfying response from plant management. Two arguments for this matter were the lack of strategic clarity and failure to acknowledge maintenance management as a separate discipline of strategic importance.

With respect to the supply chain strategy plant management was asked to translate the business line strategy into their vision of the maintenance function. As a result the management team required a reliable production process for minimal costs. This statement emphasizes the importance of a solid selection method of the most costly assets in an identical sample or crucial production position. We advised to perform an extensive and coherent study to define the relation between budget and criticality. The management team has appointed a maintenance manager with among other things this task.

4.3 STRATEGIC FOCUS DECISION RESULTS

In this respect we discuss the three categories of Table 1.

Concerning category 1, the reliability management, what is missing is mainly a good set of indicators for machines, allowing management to make appropriate decisions. In terms of strategy, design of reliability management system plays an important role. The system can as be simple one collecting mean time between failures (MTBF) or mean time to repair (MTTR), or detecting failure patterns [29, 30], and can be enriched by reliability block diagrams, reliability program, or can be further extended to full scale reliability centered maintenance [6, 31]. For the case under study we propose Reliability Centered Maintenance data gathering system to be adopted internally in order to monitor the performance of 3PMSP for the most critical assets. Obviously, such system could be made available by the 3PMSP. However, in its absence, the strategic imperative is that the company places its primary focus on developing this system since most of the data is already collected one way or the other.

Concerning category 2, the maintenance operations, two decisions must be made. First a decision should be made by grouping criticality rating levels and identifying the repair response: corrective, preventive and predictive [6, 13]. The corrective and preventive maintenance can be considered “non-core” and are provided by the current 3PMSP; predictive maintenance is of strategic importance and should be controlled more by the plant management. Second, the pool of improvement suggestions specified by the engineers should be classified more effectively and used to remove the sources of unreliability.

Concerning category 3, the maintenance related purchasing, three actions are recognized for further improvements: change of contracting process, the clear division of purchasing roles (3PMSP vs. the plant), and introduction of capital replacement models. First the contracting types are split for maintenance service by in-house or outsourcing. Next the outsourcing is split by levels of service provision: customer service, customer satisfaction or customer success levels [32]. Respectively, each level is more intensive cooperation between the production and maintenance function. This makes customer service the basic treatment; customer success the strategic treatment. Second, the purchasing types can be split as widely recognized by purchasers around the world [24]. The purchasing portfolio by Kraljic [21] contains four strategies: leverage, strategic, non-critical and bottleneck. The non-critical purchase type can easily be delegated to the 3PMSP. The capital replacement decisions often booked as capital expenses and require demonstrating savings in operational expenses. An adequate quantitative procedure to reflect the characteristics of the plant environment should be developed in order to justify the degree of improvements when a major portion of existing infrastructure is replaced.

5. CONCLUSIONS

While several considerations for a company to develop the best maintenance manufacturing strategy are discussed here, there are many others that cannot be covered in this paper due to restrictions in time and space. Needless to say, the development of maintenance strategy is a multidisciplinary work, the decision to outsource any major function, such as maintenance, is not one that should be taken lightly, and careful consideration of all major issues is vital, if the transition to contracted maintenance is to be smooth and satisfactory to both parties. This paper reported an approach to assist management in such a decision making process in a proactive manner. The approach is applied to a Dutch lube plant of a large oil company and received management acceptance.

This approach builds on specific input collected via interview and questionnaires; indicating the opportunities and challenges. The approach is the initial answer for plant management where to start. It shows four main activities: first plant management needs to highlight the opportunities and challenges, and then select the critical equipment. Next they should formalize alternative ways to address the opportunities and challenges. Subsequently they can select the appropriate strategy for every asset, based on knowledge and experience.
The study resulted in some changes in the plant management practice. They recognized that criticality tools could be adopted in terms of techniques, which assist measuring risk on the asset and plant levels. Moreover, the company introduced learning programs through cooperation between maintenance and floor operators. Furthermore, a strategic maintenance representative was appointed to supervise the interests of the company to the maintenance subcontractor and further investigate the management prioritization for the equipment maintenance.

There are some limitations to this study. First of all, the approach remains theoretical. The theories have been studied according to the case requirements; and therefore are not refined for other applications. Moreover, an implementation plan must be developed by plant management once the maintenance strategy focus is known. Second, not all the tools described in the paper are needed and sometimes other tools must be utilized. Third, the research methodology in this paper is exploratory. The weakness of exploration has effect on building a solid approach.

The approach can further be enriched by measuring the 3PMS effectiveness, by establishing an appropriate contract payment structure and linking it to the performance, and by taking the long term inter-organizational issues into the design of appropriate maintenance strategy.

REFERENCES


**APPENDIX A**

<table>
<thead>
<tr>
<th>Number</th>
<th>Variable</th>
<th>Description</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age of</td>
<td>The older; the more critical.</td>
<td>Category 5-scale Field study</td>
</tr>
<tr>
<td>2</td>
<td>Substitute</td>
<td>The more conventional the approach to substitute a defect; the less critical.</td>
<td>Category 5-scale Field study</td>
</tr>
<tr>
<td>3</td>
<td>Production Consequence Factor</td>
<td>Internal documented criteria values</td>
<td>Category 4- scale Questionnaire</td>
</tr>
<tr>
<td>4</td>
<td>Safety Factor</td>
<td>Internal documented criteria values</td>
<td>Category 4- scale Questionnaire</td>
</tr>
<tr>
<td>5</td>
<td>Financial Factor</td>
<td>Internal documented criteria values</td>
<td>Category 4- scale Questionnaire</td>
</tr>
<tr>
<td>6</td>
<td>Manufacturing, Construction and Installation Complexity</td>
<td>Internal documented criteria values</td>
<td>Category 4- scale Questionnaire</td>
</tr>
<tr>
<td>7</td>
<td>Design Factor</td>
<td>Internal documented criteria values</td>
<td>Category 4- scale Questionnaire</td>
</tr>
<tr>
<td>8</td>
<td>Usage Rate</td>
<td>The more is used; the more critical.</td>
<td>Category 5-scale Field study</td>
</tr>
<tr>
<td>9</td>
<td>Top 5 of MRB</td>
<td>The more it appears in the Top 5 of the monthly maintenance report, the more critical.</td>
<td>Category 5- scale Field Study</td>
</tr>
<tr>
<td>10</td>
<td>Access Diffculty</td>
<td>The better the access to, the less critical.</td>
<td>Interval 5- scale Questionnaire</td>
</tr>
<tr>
<td>11</td>
<td>Severity</td>
<td>The severe it is rated; the more critical.</td>
<td>Interval 5- scale Questionnaire</td>
</tr>
<tr>
<td>12</td>
<td>Occurrence</td>
<td>The more occurrence of failure on this ; the more critical.</td>
<td>Interval 5- scale Questionnaire</td>
</tr>
<tr>
<td>13</td>
<td>Detectability</td>
<td>The better a failure or cause is detected; the less critical.</td>
<td>Interval 5- scale Questionnaire</td>
</tr>
</tbody>
</table>
Designing a Network Protocol for Control Systems based on the Hypertext Markup Language (HTML)

Anwar Bashir, Farhad Nabhani, Terry Shaw

*University of Teesside
School of Science & Technology
Middlesbrough
Cleveland
TS1 3BA

Abstract

The current state-of-the-art information about protocol design in control systems is available from Common Request Broker Architecture (CORBA) and Connected Open Building Automation (COBA). The development of these systems is on-going however to date their impact has been limited. Perhaps part of the reason for this is their complexity in implementation, elements such as ORBS and IDL are not easily understood or implementable. This paper presents a new approach, based on the hypertext transfer protocol (http) that is specified as part of the ISO (International Organization for Standardization) 7 layer model. This paper describes a practical specification for http in terms of requirements for control systems. The specification is “server-centric”. It relies on “pre-amble” to establish application independent operation. The design implements re-usable modules that can be ‘switched’ in or out depending on the nature of the application. The protocols specification ensures that the software is compact and portable so that it can be included as part of the ISO’s network protocol at the application layer and incorporated as a standard network interface on control systems devices.

1. Introduction

The protocol that introduced internet technology and possibly the use of computers into the mass-market was the development of the hypertext transfer protocol (http) [1]. Essentially this protocol enables a web-browser application running on a computer with a screen to send an http request over a network. A web-server application intercepts the request and responds. The response will generally be in the form of information that is ‘tagged’ and sent back to the client. The web-browser application will receive the information, interpret the tags and present the information using the web-browser application on a screen [2] [3] [4]. Figure 1 illustrates the basic operations.

Figure 1: Basic operations of http

The job of the web-server is to accept http requests, locate the information and send it back to the computer that originated the request. However, there is a basic design limitation in this approach. The information that has been requested is effectively static, i.e. the information cannot easily change. In order to resolve this programmers have developed additional tools such as: CGI scripts (Common Gateway Interface, commonly, but not necessarily written in PERL); ASP (Active Server Pages; JSP (Java Server Pages) and PHP (HyperText Preprocessor). Often operating alongside database applications such as MySQL or SQL Server these technologies can enable the response to the client web-browser to be dynamic [5] [6].

In control systems different communication specifications need to be considered. Instead of the http request sending information back for reading, the information is received by the server to perform some control function. Another consideration is that rather than a person sending an http request via a personal computer some automated request is generated and sent via a microprocessor controller. Figure 2 depicts how a
microcontroller can send a http request and receive and instruction to perform some function by reply. In addition to responding the server can log each request.

The format of the response that is read by the microcontroller requires some consideration. An html response is generally an html page that contains header data. The header data needs to be ignored (contains browser version details etc…). The response data needs to be located. In earlier papers the authors specified that the response data should be contained within html table tags, i.e., `<table name of user</table>`. A program on the client simply searched thorough the data until a table tag was found and then stored the contents in a variable.

A later specification uses the eXtensible Markup Language (XML). This is a general purpose markup language developed by W3C (World Wide Web Consortium). It is a reduced version of the Standard Generalized Markup Language (SGML). It allows designers to create their own custom tags, enabling the definition, transmission, validation and interpretation of data between applications and between organizations. This technology requires server side XML support, however using XML allows simpler parsing, displaying and transforming of data by the client program.

The most recent specification is SOAP (Simple Object Access Protocol). This is a lightweight XML based messaging protocol used to exchange information in a decentralized, distributed environment. It consists of three parts; an envelope that defines a framework for describing what is in a message and how to process it, a set of rules for expressing instances of application-defined data types, and a convention for representing remote procedure calls and responses. SOAP messages are independent of any operating system or protocol and may be transported using a variety of Internet protocols, including SMTP, MIME, and HTTP.

2. **ISO NETWORK LAYERS**

The network transmission of data uses the TCP/IP protocol [7]. Importantly, Jon Postel wrote a document in August 1977 promoting the use of layers [8]. Layering divides protocol design into a number of smaller parts, each of which accomplishes a particular sub-task, and interacts with other parts of the protocol only in a small number of well-defined ways. Layering allows the parts of a protocol to be designed and tested without a combinatorial explosion of cases, keeping each design relatively simple. The International Organization for Standardization (ISO) developed the Open Systems Interconnection (OSI) architectural model. The model is divided into seven layers, with layers 1 to 4 called the lower layers and 5 to 7 called the upper layers. The idea of layers of data has been developed into that of datagram encapsulation. Most data originates within the higher layers of the OSI model. The protocols at these layers pass the data down to lower layers for transmission, usually in the form of discrete messages. Upon receipt, each lower level protocol takes the entire contents of the message received and encapsulates it into its own message format, adding a header and possibly a footer that contain important network control information. Data is passed to IP typically from one of the two main transport layer protocols; the Transmission Control Protocol (TCP) or the User Data Protocol (UDP). This data is encapsulated into the body of an IP message, usually called an IP datagram or IP packet. Figure 3 depicts the entire process.

The upper layer message is packaged into a TCP or UDP message. This then becomes the payload of an IP datagram; shown here with only one header (it can be more complex than this). The IP datagram is then passed to layer 2, where it is encapsulated into some sort of local area network (LAN), wide area network (WAN), or wireless LAN (WLAN) frame and then converted to bits and transmitted at the physical layer.
IP addresses are used to facilitate the routing of datagram’s in an IP network. This is made possible because of the way that IP addresses are structured and how the structure is interpreted by network routers. In IPv4 addressing the structure is 32 bits long and contains two components; the Network ID and the HostID.

- **Network Identifier (Network ID).** A certain number of bits, starting from the left-most bit, are used to identify the network where the host or other network interface is located.

- **Host Identifier (Host ID).** The remaining bits are used by the host on the network.

A typical host on a TCP/IP has many different application processes running concurrently. Each process generates data that it sends to either TCP or UDP, which then passes it to IP for transmission. The IP layer sends out this multiplexed stream of datagram’s to various datagram’s to various destinations. Simultaneously, each device’s IP layer is receiving datagram’s that originated in numerous application processes on other hosts. The datagram’s need to be de-multiplexed so that they end up at the correct process. In UDP and TCP messages two addressing fields appear: a source port and a destination port. They identify the originating process on the source machine and the destination process on the destination machine. The TCP and UDP software fills them in before transmission, and they direct them to the correct process on the destination device. TCP and UDP port numbers are 16 bits in length. Valid port numbers can theoretically take on values 0 to 65,535, however certain port numbers are reserved for particular uses.

The relationship between client/server architecture in TCP/IP becomes evident when examining how port numbers are used. Since clients initiate application data transfers using TCP and UDP, they need to know the port number of the server process. Consequently, servers are required to use universally known port numbers. Thus, well-known and registered port numbers identify server processes. Clients that send requests use the well known or registered port number as the destination port number. In contrast, servers respond to clients; they do not initiate contact with them. Thus, the client does not need to use a reserved port number. In fact this is actually unnecessary. However, to know where to send the reply, the server must be able to identify the port number that the client is using. Client processes don’t use well-known or registered ports. Instead, each client process is assigned a temporary port number for its use. This is commonly called an ephemeral port number.

So, the overall identification of an application process actually uses the combination of the IP address of the host it runs on – or the network interface over which it is talking, to be more precise – and the port number that has been assigned to it. This combined address is called a socket. Sockets are specified using the notation <IP Address>:<Port Number>. For example, if you have a website running on IP address 41.199.222.3, the socket corresponding to the HTTP server for that site would be 41.199.222.3:80 (Port 80 is the standard assignment for HTTP servers).
3. APPLICATION LAYER TECHNOLOGIES

The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. Their involvement in the XML standard and to some extent SOAP is achieving some success. The Universal Plug and Play protocol (UPnP) is an industry initiative designed to enable simple and robust connectivity among consumer electronics, intelligent appliances and mobile devices from many different vendors.

XML – Extensible Markup Language is a general-purpose markup language developed by W3C. It is a reduced version of the Standard Generalized Markup Language (SGML). It is especially useful for preparing web documents. It allows designers to create their own custom tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

SOAP – Simple Object Access Protocol is a lightweight XML based messaging protocol used to exchange information in a decentralized, distributed environment [9] [10]. It consists of three parts: an envelope that defines a framework for describing what is in a message and how to process it, a set of encoding rules for expressing instances of application-defined data types, and a convention for representing remote procedure calls and responses. SOAP messages are independent of any operating system or protocol and may be transported using a variety of Internet protocols, including SMTP, MIME, and HTTP.

UPnP is built upon existing protocols [11] and formats including HTTP, XML, SOAP, the Document Object Model, and IP multicast. These elements are used together to define a framework to enable the definition, discovery, and control of network devices. In its current form the authors would assert that this protocol has limited application in control systems. The research being undertaken may however extend UPnP the functionality of this protocol to incorporate the required functionality.

4. USING HTTP TO DEVELOP CONTROLLER BASED APPLICATIONS

The aim of the research is to utilize the http protocol and specify an application layer protocol so that it can be incorporated into any networked control device thereby enabling standardization across a vast number of control devices. Once plugged into a network the protocol would enable the device to automatically perform appropriate functions without the need for bespoke programming.

The protocol development began with a basic prototype. A microcontroller was programmed to send an http request to a server. The http request would submit a user identity. The server receives the request and responds with the name of the users identity as affirmation the user is known or with text “user not known”. Figure 4 depicts the following operations;

“The microcontroller 1 reads an identity from a reader (not shown). An http request 2 containing the identity is sent to server 3. The request instantiates a database query 4 which obtains a name match for the identity. The server 4 then returns an html page that contains 5 that contains the name or a blank. If a web-browser received the information it would display the contents on the screen as “Anwar”, however is this case the microcontroller receives the data and, ignoring the header information searches for data between an exclamation marks (later versions search for table tags and the latest version searches for XML markers). It then stores the name in a variable.”
A practical implementation of these operations may be an access control system; an identification device (swipe card reader, proximity card reader or some other) is connected to a microcontroller. When a card is presented it is read and the data is sent via an http request to a server. The server intercepts the request and responds. The response may be ‘name of user’ or ‘user not found’. This information will have been generated by some server-side application data processing. The ‘browser’ receives the response. The control is now dependant on the response received. If the response was ‘user not found’ then the door actuator remains closed, if the response was ‘name of user’ then the door actuator is opened.

Initial configuration and testing of the prototype suggests that the design of the protocol should ensure four functional principals are addressed:

1. **Preamble (Plug & Play)**
   The microcontroller should be application independent until a preamble establishes the exact operational requirements. The software modules should be loaded as part of the pre-amble. The pre-amble should be server-centric.

2. **Server-Centric**
   Microcontrollers can be powerful devices with storage capability. It can be tempting therefore to utilize their facilities to transfer and store data locally. Minimal information should be sent from the microcontroller. Data-processing should be conducted on the server. The server should respond to the microcontroller only with information that is absolutely necessary to perform some function.

3. **Network reliability**
   The protocol design requires network reliability. A process to ensure network connectivity should be in place and design working parameters must be established. Exception handling for outages or operations outside these parameters must be programmed.

4. **Re-usable modules**
   In considering a specific requirement the software should consider re-use by other applications. For example, the device that is used to read an identity could change (swipe card or RFID or biometric device) but the basic http request can remain the same.

5. **A MULTIFUNCTIONAL PROTOCOL SPECIFICATION FOR CONTROLLERS**
   The basic communications of the multifunctional protocol is depicted in Figure 5. The client required network resilience and exception handing in case of network outage or operations outside acceptable parameters. This is a continuous process. The client is basically a ‘dumb’ microcontroller with network capability. Upon boot-up it acquires a network connection and communications with a server. It informs the
Designing a Network Protocol for Control Systems based on HTML

The server about its capability. The server responds with a list of applications which it knows can be supported. The client chooses an application from the list and requests an upload of one of the applications (Note: the application is likely sharing modules with many different applications). After uploading the client runs the application. The client only uploads the application once, only after a reset or software update availability will the client re-load.

Figure 5: Basic protocol communications

The protocol design enables clients with different control requirements to be configured from a pool of server applications. Figure 6 depicts a number of applications that have been designed and deployed at the University of Teesside. The ‘Heidy’ machine controller does not require the access control module (depicted as the key icon) whilst the other controllers have uploaded this module as part of the application. The data-logging allows web-based applications to interface with applications.

Figure 6: Multifunctional control system
The microcontroller software was broken down into separate libraries, see Figure 7. The main program contains the main program loop. The program selects the application that the microcontroller is to perform. This is currently a manual selection that is made using the terminal interface, but could be automated.

- TCP/IP library has been pre-built by the microcontroller manufacturers and enables network communications.
- FTP Server library is based on a pre-built library but has been amended so as to receive print files and direct them to the microcontroller’s parallel port.
- Input/Output library has been built to initialize the microcontrollers input and output ports. It configures the ports in accordance with the applications requirements. The library contains the programs that are able to communicate with a swipe card reader, iButton, RFID and other devices.
- The Request/Reply library contains the HTTP requests and programs that receive the returned data. The data received is assumed to have a common structure; the data of interest would be delimited by HTML table markers.
- The copier, door access, heidy and printer control libraries contain the programs that are specific to each application.

The main loop, together with these libraries, enables the microcontroller to communicate with multiple devices and create applications that can share common information. An examination of multi-functionality can then reveal additional common structures and provide the basis for further research into how a range of diverse applications can share data structures and programs.

REFERENCES

http://www.w3.org/Proposal.html.
RFID-Enabled Real-Time Mass-Customized Production Planning and Scheduling

George Q. Huang, Mike J. Fang, Haili Lu
Department of Industrial and Manufacturing Systems Engineering
The University of Hong Kong
Hong Kong, PR China

Q.Y. Dai
School of Information Engineering
Guangdong University of Technology, PR China

WN Liu
College of Computer Science
Chongqing University, China

Stephen T. Newman
Department of Mechanical Engineering
Bath University, UK

ABSTRACT
This paper discusses the use of RFID technologies to enable the shopfloor visibility and reduce uncertainties in the planning and scheduling for manufacturing engineered-to-order products of wide variety and small quantity. Several contributions are significant. Firstly, wireless devices such as RFID and Wi-Fi are deployed into value-adding points to create smart objects in a wireless manufacturing (WM) environment for the collection and synchronization of real-time shopfloor data. Secondly, the medium-term planning horizon is divided into multiple time periods. Within each period, disturbances and changes are tracked and traced by RFID-enabled real-time events. The third contribution is a bilevel model for planning and scheduling decisions, together with the novel concept of order progress kanbans as a decision coordination mechanism between the two levels. Finally, an event-driven real-time scheduling model is established for a hybrid flowshop based on a collaborating company’s workshop. A backward method is developed for solving this problem with near-life complexity.

Index Terms—RFID / Auto ID, Wireless Manufacturing, Mass Customized Production, Manufacturing Execution System (MES)
1. INTRODUCTION

Mass customization (MC) is aimed at producing a wide variety of products that are engineered, customized and/or personalized in large scale to satisfy diverse customer requirements within intended market segments, still benefiting from the scale of economy as in mass production[1]. MC can be achieved through efforts at different levels. First, products must be designed following Design for Mass Customization (DFMC) principles[2]. The platform approach has been widely considered an effective way of creating product designs suitable for MC[3]. Second, supply chain and manufacturing processes must be adequately flexible and reconfigurable to accommodate the necessary variety at the levels of end products, components and materials. Huang et al.[4] and Huang and Qu[5] present methods for simultaneously configuring mass-customized products and corresponding supply chains & manufacturing processes. Koren et al.[6] focus on developing flexible and reconfigurable machining centres. Finally, production planning and scheduling (PPS) decisions must be made within given shopfloors to produce Mass-Customized components and final product assemblies. Tu[7] presents production planning and scheduling methods for one-of-a-kind products.

This study uses the term MCP (Mass-Customized Production) to define the mode of workshop for producing mass-customized products. The work has been motivated by several industrial collaborators involved in discrete manufacturing of a large variety of component module options and product variants of small or one-of-a-kind (OKP) quantities. Products include molds and dies, lifts and industrial equipment.

For convenience of reference, this paper will use mould and die (MD) manufacturing as an example to discuss MCP planning and scheduling (MCPSS). MD products are OKP and engineered-to-order (ETO) products. Only one or several product is typically ordered and manufacturing is non-repetitive. Key components of different MD products are unique and must be made ready before the final assembly with little interchangeability. The arrival of the large number of customer orders and operation times are highly uncertain. This is further compounded by uncertain disturbances such as frequent engineering changes and emergency orders. They disrupt production plans and schedules. As a result, orders cannot be fulfilled on a timely basis and bottlenecks develop at critical machining centers. Fire-fighting is common between the order/project manager whose primary objective is to deliver the products on time, and shopfloor manager whose primary objective is to maximize productivity and machine utilization.

The long-term of this research is to develop a solution to overcome MCPSS difficulties mentioned above. Several measures are taken to deal with uncertainties. Firstly, advanced wireless devices such as RFID and Wi-Fi are deployed into value-adding points to create so-called wireless manufacturing environments for the collection and synchronization of real-time shopfloor data. Secondly, the medium-term planning time horizon is divided into multiple shorter periods. Within each period, disturbances and changes are tracked and traced by RFID-enabled real-time events and uncertainty is minimized. The third measure is the concept of order or project progress kanbans which provide a coordination mechanism. Shopfloor conditions such as machine loading and job due dates are updated and derived through RFID-enabled real-time events during the concerned time period for the forthcoming period to be planned and scheduled.

This paper is structured as follows. Section 2 reviews the related literature on real-time wireless manufacturing. Section 3 presents how RFID technologies are deployed in a typical MCP shopfloor. Section 4 discusses a bilevel model for MCP planning and scheduling. Section 5 gives the detail on the lower-level event-driven real-time shopfloor scheduling. The paper is concluded in Section 6.

2. LITERATURE REVIEW

Related work is categorized according to the following list:

- MD manufacturing as OKP planning & scheduling.
- Hierarchical Production Planning & Scheduling (HPPS).
- Project and production scheduling under uncertainties.
- “Push and Pull” production planning & scheduling.
- Real-time shopfloor data capturing.
- RFID-enabled real-time distributed dynamic scheduling.
Taking mould and die (MD) products as an example, the order size is usually small in single digit. Therefore, MD manufacturing is OKP. OKP planning and scheduling is complex. Tu and his co-workers have conducted a series of research in the area [7, 8]. Most references on MD PPS as OKP are based on deterministic models without considering uncertain disturbances [9, 10]. Lee et al. [11] consider the order acceptance problem incorporating machine loading information. Choi et al. [12] develop a framework for die manufacturing management. Iyama et al. [13] give a case study and develops dispatching rules to explicitly consider coordination of the completion time of different mold parts for just in time assembly.

HPPS (Hierarchical Production Planning and Scheduling) model decomposes the decision process of production planning and scheduling into subproblems with different levels of decision variables across different time horizons. Early research on this topic includes [14, 15]. McKay reviews the relevant HPPS literature [16]. Qiu and Burch [17] formulate the problem similar to bilevel model but solved the problem differently.

The literature on project scheduling under uncertainty is scarce. A typical treatment is to consider the project network as deterministic while some of activity parameters as stochastic [17]. A more general consideration is to treat project networks as stochastic[18]. Most references on scheduling problems under uncertainty are for production scheduling. Aytug et al.[19] provide a comprehensive review of the literature, covering reactive, predictive, stochastic and fuzzy scheduling. Sabuncuoglu and Bayiz [20] and Vieira et al. [21] discuss reactive scheduling.

Benton and Shin[22] compare and contrast MRP (Material Requirement Planning) and JIT (Just In Time) as “Push” and “Pull” representatives. Their integration is discussed to use their strengths while avoid their limitations. A typical integration is to use JIT kanbans together with MRP system. Deleersnyder et al.[23] conduct comparative analysis of push, pull and push-pull production control with a conclusion that the push-pull mixture system is more responsive to changes but simpler to use. Hirakawa et al.[24] examine the capability of the push-pull system in dealing with changes in production and inventories. However, kanbans mechanism is more suitable for repetitive manufacturing. Its use for OKP is rare in the literature.

The need for automatic object identification (Auto ID) in manufacturing shop-floors has been long recognized in the field of Computer Integrated Manufacturing (CIM) [25,26]. The dropping cost of RFID technology (tags and readers) have motivated worldwide sporadic piloting efforts across different product sectors ranging from garment, electronic, mechanical, aerospace and automotive products. Huang et al. [27] provide an in-depth review of the recent developments of RFID manufacturing applications.

Real-time scheduling methods are reviewed and discussed by various researchers[17,20,28-33]. However, most studies have not been conducted in the context of RFID/Auto-ID-enabled real-time events. Paper by Kohn et al.[34] is a piece of precious early work in addressing repair-control of manufacturing systems using real-time RFID information.

3. WIRELESS MCP SHOPFLOOR

Fig. 1 shows the wireless manufacturing workshop as deployed in one collaborating company. The questions like what are tagged, where are RFID readers deployed and how do they communicate with each other are answered in the diagram.

The deployment follows the roadmap of RFID-enabled Wireless Manufacturing Infrastructure as set out in Fig. 2. This WM platform achieves real-time and seamless dual-way connectivity and interoperability between application systems at enterprise, shop floor, work-cell and device levels. Accordingly, the following four core components:

- Shop floor Gateway: The Shop-floor Gateway (SF-Gateway) is at the center of the overall WM platform. Its architecture is based on the Service-Oriented Architecture (SOA). SF-Gateway includes three main components, i.e. workflow management tools, MS-UDDI and agent-based manufacturing services management tools, which will be introduced in next chapter in details.

- Work-cell gateways: Work-cell Gateway (WC-Gateway) acts as a server that hosts and connects all RFID-enabled smart objects of the corresponding work-cell. A WC-Gateway has a hardware hub and a suite of software systems. The hub acts as a server that hosts and connects all RFID-enabled smart objects of the corresponding work-cell. Smart objects are represented as software agents in the WC-Gateway operating system within which they are “universal plug and play (UPnP)” and interoperable.
RFID-enabled smart objects: Smart objects are those physical manufacturing objects that are made “smart” by equipping them with RFID devices. Those with RFID readers are active smart objects. Those with RFID tags are passive smart objects. Smart objects interact with each other through wired and/or wireless connections, creating what is called an intelligent ambience. In addition, smart objects are also equipped with their specific operational logics, and with data memory and processing functions. Therefore, smart objects are able to sense, reason, act/react/interact in the intelligent ambience community.

![Fig. 1 RFID-enabled manufacturing workshop environment](image1)

![Fig. 2 RFID-enabled Wireless Manufacturing Infrastructure](image2)

**Notes on RFID deployment**
- All machines are equipped with RFID readers which are networked through wires or wirelessly.
- All operators have RFID-enabled staff cards.
- All job components (or their carrying containers / trays) are RFID tagged.
- RFID tag is read when a job enters and leaves a machine – triggering event chain.
- Information Infrastructure is implemented to deal with real-time RFID events and information.

**How to make decisions under RFID-enabled events?**
- How to deal with frequent changes and uncertainties?
- How to balance competing objectives of project and production managers?

Fig. 1 RFID-enabled manufacturing workshop environment

Fig. 2 RFID-enabled Wireless Manufacturing Infrastructure
4. Multiperiod Bilevel MCP Planning and Scheduling

In this work, MCPPS decisions in a hybrid flowshop are optimized in a multi-period bilevel model as shown in Fig.3:

- Assigning due dates by determining reliable and quick throughput times can be promised;
- Releasing orders by determining when a manufacturing unit receives a work order; and
- Sequencing orders by determining the priority of work orders at each machine stage.

Fig. 3 Multi-period bilevel MCP planning and scheduling

A. Multi-period of Planning Horizon to Deal with Uncertainties

Order releasing and assignment decisions are solved over a longer planning horizon T which is divided into multiple time periods of equal-length t. The current time instant is assumed at the end of period te, which is also the time instant for an MCPPS execution. The scope of the current MCPPS run includes every time period t, such that te ≤ t ≤ te+T. The first two decisions are optimized over macro T. The sequencing is accomplished in a micro period t through a pull mechanism. Multi-period model dynamically handles uncertain changes in operation times and machine availability. Its purpose is to mitigate the discrepancy between the forecast-driven push activities and job-driven pull activities across the entire flowshop. As a result, planning can match execution mechanisms on every push-pull boundary across stages to improve performance.

B. Shopfloor (Lower-Level) Subproblem

In our collaborating company, order releasing and sequencing decisions are made by two independent decision makers: order / project manager and workshop manager. They usually have competing objectives. In order to achieve the optimal, a bilevel model is formulated in this study.

The shopfloor manager is the lower-level follower who determines the sequences of jobs that are already released to queues (called job pools) at individual stages within the current time period t. The objective is to minimize the makespan, also maximizing machine utilization. Let \( C_j \) denote the completion time of job j. This lower-level subproblem is defined as

\[
\text{Min } C_{\text{max}} = \max\{ C_1, C_2, \ldots, C_f \}
\]

s.t. Constraints
C. PROJECT (Upper-Level) SUBPROBLEM

An order project manager is the upper-level leader who determines when jobs (product components) should enter and/or leave job pools of which stages of the flowshop at the end of t. Which objective the order manager should use depends on the specific flowshop situation. For example, the minimization of the average flow time of all the projects (orders) not only ensures that a project order is fulfilled as quickly as possible by completing all its component jobs as shortly as possible (within as short period available for assembly), but also minimizes WIP inventories. Let \( f_t \) denote flow time for order \( l \) and \( L \) be the total number of orders being planned. Therefore, this subproblem is defined as

\[
\min \frac{1}{L} \sum_{l=1}^{L} f_t \\
\text{s.t. Constraints.}
\]

D. BILEVEL PROGRAMMING MODEL

Upper-level and lower-level subproblems are combined to formulate the following bilevel model:

\[
\min \frac{1}{L} \sum_{l=1}^{L} f_t \\
\text{s.t. Constraints}
\]

\[
\min C_{\text{max}} = \max\{ C_1, C_2, \ldots, C_J \};
\]

\text{s.t. Constraints}

E. ORDER PROGRESS KANBANS AS COORDINATION MECHANISM

Some jobs may be related to each other. For example, they may belong to the same customer order. Their release and sequencing at the two levels must be coordinated. This research introduces a concept called Progress Kanbans. This mechanism maintains the status of a customer order in terms of constituent job orders. It is typically used for tracking and tracing internal production jobs and external outsourcing jobs that constitute an order. Two important functions are to visualize and evaluate the progress of the chosen order and its jobs. Actual operation times are recorded and used to dynamically evaluate and update due times of individual jobs from the perspective of fulfilling the whole order. More detail is given in Huang et al (2009).

5. EVENT-DRIVEN REAL-TIME SCHEDULING

Based on the case of our industrial collaborators, components of the same MD product can be produced in parallel without precedence relationships. All components must be processed through \( K \) stages and stage \( k \) consists of \( P_k \) homogeneous machines. Because all components of the same product are assumed to share the same flowshop, components are treated as jobs. Each stage \( k \) provides a job pool of \( b_k \) size, as a queue of work in progress (WIP) oncoming jobs. Stages do not have outgoing pools. When a machine at stage \( k \) becomes available, the first job in the pool is assigned to it. When the job is completed, it is immediately sent to the job pool of next stage \((k+1)\). Job pools are re-sequenced for every few period. This repeats until the job is completed at stage \( K \), as shown in Fig. 4. This section discusses real-time dynamic sequencing problem.

A. EVENT-DRIVEN BACKWARD SEQUENCING

The lower-level MCPPS problem is a problem of real-time sequencing of job pools of all stages of the hybrid flowshop whenever major change events such as machine breakdowns, new arrivals / departures of jobs in a stage
pool are detected within a time interval at RFID-enabled smart objects. The use of this time-interval instead of a single event improves the schedule stability while allowing ample time for computation. The length of the time interval is also small enough to avoid the situation where stage job pools become empty.

The objective of sequencing is to determine the optimal sequence (priorities) of jobs in the job pool of all stages. This priority informs which job an operator should pick using his/her staff card. In a hybrid flowshop, the sequencing decisions can be made stage by stage from the back (last stage) with relative independence. Fig. 5 shows the flowchart of the backward “pull” sequencing method for a hybrid flowshop. The first and last stage problems require special considerations and all intermediate stage problems are the same.

**B. SEQUENCING AT THE LAST STAGE**

A backward propagation strategy is proposed in this work. The strategy starts with considering the last stage - Stage K. Two criteria should be considered:

1. Jobs in the pool must be accomplished on time. This criterion implies that the deviation from the target due time for all jobs in the pool must be minimized. That is, the total earliness and tardiness must be minimized.

2. Jobs in the pool must be fulfilled as fast as possible. This criterion implies the makespan minimization which maximizes the machine utilization.

This multi-criteria optimization problem can be simplified as follows. The minimum makespan is found first. Multiple solutions often exist. This value is then used as a constraint. The on-time delivery is used as the objective function again to identify the best sequence that satisfies both criteria.
Let $C_{k,j}$ denotes the completion time of job $j$ on stage $K$. Let $d_{k,j}$ denotes the due time of job $j$ on stage $k$ which is determined by the upper-level, and $|C_{k,j} - d_{k,j}|$ is the earliness or tardiness of job $j$ on stage $K$. Makespan and on-time delivery are combined as the objective function weighted by $\alpha$ and $\beta$ respectively:

$$\text{Min} \quad C_{\text{max}} = \alpha \max\{C_{1,1}, C_{1,2}, ..., C_{i,j}\} + \beta \sum_{j=1}^{n} |C_{i,j} - d_{i,j}|$$

**C. Sequencing at an Intermediate Stage**

The sequence of job pool of an intermediate stage $i$ is optimized by the same criteria used for the last stage above. The on-time delivery criterion is measured in the same way above. The makespan, however, requires some discussion.

This sequencing problem at stage $i$ is similar to that of a two-stage (i.e. Stages $i$ and $(i+1)$) hybrid flowshop scheduling problem. Let $C_{i,j}$ and $C_{i+1,j}$ denote the completion time of job $j$ on stage $i$ and stage $(i+1)$ respectively, and $d_{i,j}$ is the due time of job $j$ on stage $i$. The problem weighted by $\alpha$ and $\beta$ respectively becomes:

$$\text{Min} \quad C_{i,max} = \alpha \max\{C_{i,1}, C_{i,2}, ..., C_{i,j}\} + \beta \sum_{j=1}^{n} |C_{i,j} - d_{i,j}|$$

However, this problem is different from two stage hybrid scheduling problem in the sense that the sequence of jobs in the pool of Stage $(i+1)$ is already fixed. Nevertheless, this sequence affects the sequence of jobs in Stage $i$’s pool. The minimization of $C_{i+1,j}$ maximizes the machine utilization at both stages $i$ and $(i+1)$, that is, to minimize their idle times.

**D. Scheduling at the First Stage**

The first stage problem is basically the same as any intermediate stage problem. An exception of Stage 1 problem is that its job pool must be fully filled for each period $t$. This filling process is actually what is called job release. Job releasing belongs to the production planner who maintains an order or project pool. The order pool is sequenced in the upper-level MCPPS problem. If there are $n$ vacant spaces in the Stage 1’s job pool, the first $n$ jobs are released from the order pool to the Stage 1 job pool which is sequenced in the same way as an intermediate stage.

**E. Solution Algorithm**

Sequencing problems of individual stages can be solved separately from the back. Each stage problem is very straightforward to solve. Genetic Algorithm (GA) is used for its efficiency in finding near-optimal solutions quickly. Due to limited space, the discussion of GA is omitted here.

**6. Concluding Remarks**

This research has aimed itself at capitalizing the benefits of RFID-enabled wireless manufacturing (WM) technology for real-time planning and scheduling for mass-customized production (MCPPS) of wide variety of engineered-to-order products of small quantity. With the support of the WM technology, this complex MCPPS problem has been decomposed and simplified into two levels of decisions which are coordinated by an RFID-enabled order/project progress kanban mechanism. The paper has focused on the lower-level dynamic sequencing problem of job pools of stages in a hybrid flowshop. Within a given short time period, this sequencing problem is formulated as individual decision models for stages and solved from backward independently stage by stage. The problem at each stage is solved as an event-driven real-time sequencing of its job pool within a short time period $t$. This treatment significantly improves the computational efficiency required for real-time decisions.

Further research and development are necessary if the proposed MCP planning and scheduling model is used practically in the RFID-enabled wireless manufacturing shop-floors. Firstly, the two levels of decisions should be truly considered in a bilevel programming model. Secondly, the concept of project kanbans is needed as a coordination mechanism between the two levels and among related jobs. Thirdly, jobs may not go through all
stages. That is, some job may be finished at earlier stages and by-pass several intermediate stages. Fourthly, machines at the same stage may have different capabilities and the same job may need different operation times on different machines even at the same stages. Finally, changeover/setup times may be sequence-related. In addition to relaxing these assumptions, it would be necessary to conduct a series of sensitivity analysis to study the impacts of macro and micro time periods (T and t) and impacts of parameters of jobs and orders.

REFERENCES


Inventory Lot-sizing and Supplier Selection with Shortage Cost in an Assembly System

Jafar Razmi* and Ghazal Assadipour

Department of Industrial Engineering, University College of Engineering
University of Tehran, P.O. Box: 11155-4563, Tehran, Iran

Abstract

Assembly systems are the base of production in great number of industries. These production systems are more vulnerable to scarce of components when supply process is uncertain; therefore one of the main tasks of the supply department is to consider the fluctuations of supplier's capacity, when ordering. This paper considers a single product case, assembled of multiple components, with multiple suppliers whose capacities' are variable through the multi-period horizon of known demand. To produce the final product, all components that can be supplied by all suppliers, should be available for assembly. Since the shutdown cost is high, the manufacturing will be resumed even though some components are not supplied, and the incomplete products either stay near the line as they need the assembly line to set up again for the assembly of the slacking component(s), or conveyed to the inventory and remain there until the lacking components are received. To find the optimal ordering policy, a mixed integer mathematical model is formulated, which helps the decision maker determine the type, amount, supplier and periods of order with regard to variability of prices, suppliers’ capacities, setup, holding and delay costs. As the model cannot be solved in an acceptable time for large scale problems a Particle Swarm Optimization (PSO) algorithm is developed.

1. Introduction

Since 1958, when Wagner and Whitin [1] proposed a dynamic programming algorithm for their single product lot-sizing problem, different extensions of their model have been developed. Hariga and Haouari [2] assumed that supplier has a random variable capacity and formulated the lot-sizing model under the EOQ framework to determine the optimal lot-size which minimizes the expected cost per unit of time. They also showed that the functional form of the optimal ordering quantity is affected by the random capacity. Another version of Wagner and Whitin’s model is presented in [3], where backlogging and start-up costs are considered. With regard to Wagner-Whitin costs they suggested a linear extended model with an exponential number of constraints. A combination of lot-sizing and supplier selection is modeled in [4]. Basnet and Leung altered the traditional lot-sizing model to a multi-product and multi-period case with multiple suppliers. Considering the limitation on suppliers’ capacities, Benson [5] extended work of [4] and determined the optimal procurement strategy with multiple capacitated suppliers. Assuming that received items might be imperfect and can be sold in lower price, Rezaei and Davoodi [6] adapted Benson’s model and developed a mixed integer programming model. Absi and Kedad-Sidhoum [7] studied a capacitated lot-sizing problem with multiple products and setup times. Assuming that demand cannot be backlogged but can be lost, shortage costs are incorporated in the model which is solved through a branch-and-cut algorithm.

In this paper we seek the optimal order policy in a single product supply chain, with multiple capacitated suppliers, through a multi period horizon of known demand. Each final product is composed of specific number of components which can be supplied by all suppliers. Each period’s demand can be satisfied by production, by stock carried over from previous periods or can be backlogged. Because of the uncertainty of supply, a product may remain incomplete while is lacking in at least one component. Incomplete products are classified into two groups: those which stay near the line, as their satisfaction requires the line to setup again, and those which can be conveyed to the incomplete inventory until their lacking components are delivered. In addition, suppliers’ transaction cost,
suppliers’ capacities, components’ purchase cost and the final product selling price are considered to be time
dependent.

The problem is formulated as a mixed integer programming model in the following section. Then section 3
describes the PSO algorithm which is developed to solve the model. In section 4, an example is presented, and
finally section 5 concludes our work.

2. PROBLEM DEFINITION

We model a supply chain with a single, but multi-component, product. The product is complete and ready for sell,
only if adequate numbers of all components are available. Otherwise two scenarios are possible: either the
incomplete product is lacking in components whose satisfactions require the line to setup again and consequently are
kept near the line, or can be conveyed to the incomplete inventory and preserved there until the lacking components
are delivered. Thus, components are categorized into three groups:

Essential components: main items which are assumed to be always available and all products, complete or
incomplete, contain them.

Process components: products which are lacking in this type of components are called Process Incomplete here.
These products are kept near the line until the lacking parts are received and need the line to setup again for
resuming the process. Process Incomplete products contain the Essential components but are lacking in Process and
probably Inventory components.

Inventory components: products which are lacking in this type of components are called Inventory Incomplete here.
These products contain Essential and Process components while are lacking in Inventory components. Inventory
Incomplete products are conveyed to the incomplete inventory and are kept there until the lacking components are
delivered.

The supply chain is considered to be multi-supplier and multi-period. All components can be supplied by all
suppliers whose capacities are limited and variable through the multi-period horizon. In addition, a time and supplier
dependant transaction cost applies for each period in which an order is placed on a supplier. The demand over the
horizon is known, which can be satisfied by production, by stock carried over from previous periods or can be
backlogged. A sample supply chain with 3 suppliers and 3 components is demonstrated in Figure 2. It is supposed
that one unit of each component is required to assemble one unit of the final complete product, so the Process
Incomplete products are lacking in the second and probably the third components, while the Inventory Incomplete
products are lacking in the third component. We assume that satisfaction of incomplete products, carried from
previous periods, is prior to production of new products.

Apart from the product dependent holding cost which is applied for each component in the inventory that is carried
across a period in the planning horizon, two other holding costs are considered: a holding cost for surplus complete
products which are ready to sell but should be stocked for the successive periods when there will be demands, and
another holding cost for Inventory Incomplete products which are kept in the incomplete inventory until the lacking
components are supplied in the next periods. In addition, when two Process Incomplete and Inventory Incomplete
products are lacking in the same component, the satisfaction of Process Incomplete takes priority over Inventory
Incomplete product.
The problem of lot-sizing and supplier selection of an assembly system with multiple components through the multi-period horizon is formulated as below:

Parameters:

- $N_i$ = number of component $i$ required to assemble one unit of finished product.
- $U_i$ = 1 if component $i$ is an Essential component, 0 otherwise.
- $V_i$ = 1 if component $i$ is a Process component, 0 otherwise.
- $W_i$ = 1 if component $i$ is an Inventory component, 0 otherwise.
- $tD_t$ = demand for complete product in period $t$.
- $ijtB$ = purchase cost of component $i$ from supplier $j$ in period $t$.
- $jtO$ = transaction cost for supplier $j$ in period $t$.
- $tS_t$ = selling price of complete product in period $t$.
- $M$ = unit manufacturing (assembly) cost.
- $tSC_t$ = shortage cost of complete product in period $t$.
- $SU$ = set up cost of Process Incomplete products.
- $iHC$ = unit holding cost of component $i$ per period.
- $HP$ = unit holding cost of complete product per period.
- $HI$ = unit holding cost of Inventory Incomplete product in incomplete inventory per period.
- $Q_i$ = storage space required by component $i$.
- $ICC$ = capacity of components’ inventory.
- $IIC$ = capacity of Inventory Incomplete products’ inventory.
- $IC$ = capacity of complete products’ inventory.
- $m$ = a large constant.

Decision variables:

- $ijtX$ = number of units of component $i$ ordered from supplier $j$ in period $t$.
- $jtY$ = 1 if an order is placed on supplier $j$ in time period $t$, 0 otherwise.
- $itR$ = inventory of component $i$, carried from period $t$ to period $t+1$.
\( H^+_t \) = inventory of complete product, carried from period \( t \) to period \( t+1 \).

\( H^-_t \) = backlog of complete product, carried from period \( t \) to period \( t+1 \).

\( A_t \) = selling amount of complete product in period \( t \).

\( F_t \) = number of complete products manufactured in period \( t \).

\( TF_t \) = total number of complete products from the beginning until period \( t \).

\( TE_t \) = total number of products containing the Essential components from the beginning until period \( t \).

\( P_t \) = number of Process Incomplete products in period \( t \).

\( I_t \) = number of Inventory Incomplete products in period \( t \).

With the above notations the problem can be formulated as:

Max \( Z = \sum_{i=1}^{T} S_i A_t - \sum_{j=1}^{T} \sum_{t=1}^{T} O_j y_{ijt} - \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{t=1}^{T} B_{ijt} X_{ijt} - \sum_{i=1}^{T} H_{iC,R} - \sum_{t=1}^{T} M.E_t - \sum_{t=1}^{T} H.P.H^+_t - \sum_{t=1}^{T} S.C.H^-_t - \sum_{t=1}^{T} S.U.P_t - \sum_{t=1}^{T} H.I.t \) \hspace{1cm} (1)

s.t.

\( U_i + V_i + W_i = 1 \quad \forall i \) \hspace{1cm} (2)

\( TF_t = \min \{ ( \sum_{k=1}^{T} \sum_{j=1}^{T} X_{ijk} ) / N_i \} \quad \forall t \) \hspace{1cm} (3)

\( F_t = TF_t - TF_{t-1} \quad \forall t \) \hspace{1cm} (4)

\( TE_t = \min \{ ( \sum_{k=1}^{T} \sum_{j=1}^{T} (X_{ijk} + m(1-U_i)) ) / N_i \} \quad \forall t \) \hspace{1cm} (5)

\( E_t = TE_t - TE_{t-1} \quad \forall t \) \hspace{1cm} (6)

\( P_t = TE_t - \min \{ ( \sum_{k=1}^{T} \sum_{j=1}^{T} X_{ijk} + m(1-V_i) ) / N_i \} \quad \forall t \) \hspace{1cm} (7)

\( I_t = \min \{ TE_t - P_t \} - \min \{ ( \sum_{k=1}^{T} \sum_{j=1}^{T} (X_{ijk} + m(1-W_i)) ) / N_i \} \quad \forall t \) \hspace{1cm} (8)

\( R_{it} = \sum_{k=1}^{T} \sum_{j=1}^{T} X_{ijk} - N_i ( \sum_{k=1}^{T} F_{ik} + P_t + I_t ) \quad \forall i,t \) \hspace{1cm} (9)

\( A_t = \min \{ ( \sum_{k=1}^{T} F_{ik} , \sum_{k=1}^{T} D_{ik} ) - \min \{ ( \sum_{k=1}^{T} F_{ik} , \sum_{k=1}^{T} D_{ik} ) \} \quad \forall t \) \hspace{1cm} (10)

\( H^+_t = \sum_{k=1}^{T} F_{ik} - \sum_{k=1}^{T} D_{ik} \quad \forall t \) \hspace{1cm} (11)

\( H^-_t = \sum_{k=1}^{T} D_{ik} - \sum_{k=1}^{T} F_{ik} \quad \forall t \) \hspace{1cm} (12)

\( X_{ijt} / N_i \leq ( \sum_{k=1}^{T} D_{ik} ) Y_{jkt} \quad \forall i,j,t \) \hspace{1cm} (13)

\( \sum_{j=1}^{T} \sum_{t=1}^{T} X_{ijt} / N_i \geq \sum_{t=1}^{T} D_t \quad \forall i \) \hspace{1cm} (14)
The objective is to maximize total profit, and consists of nine terms. First term is the total revenue of selling final products. The amount of products can be sold in each period never exceeds the unsatisfied demand up to that period. The remaining terms calculate the total cost: ordering, purchasing and holding costs of components. Fourth term refers to manufacturing or assembly cost of products, while holding and shortage costs of final products are computed in fifth and sixth. The last two terms calculate setup cost of Process Incomplete products and holding cost of Inventory Incomplete products. First constraint categorizes components into three groups of Essential, Process and Inventory components. Constraints 3 to 12 compute the necessary variables. Thirteenth constraint stipulates that an ordering cost is charged by the appropriate supplier whenever an order is placed. Constraint 14 ensures that all demands are satisfied by the end of the last period. Thus there is no obligation for the demands to be fulfilled in the period they occur, and backlog is allowed in all periods except the last one. Constraints 15 to 17 are the inventories’ capacity restrictions and constraint 18 states that suppliers have limited capacities.

The scale of the presented mixed integer non-linear model, which is determined by the number of constraints, is calculated as follows:

\[
\text{scale linearNon}(I,J,T) = (2 \times I) + (12 \times T) + (I \times T) + (3 \times I \times J \times T)
\]

It is obvious that scale grows as the number of components, suppliers and time periods increase and the situation even get worst when extra intermediate restrictions and binary variables are employed to make the model linear. Thus a PSO algorithm is proposed to solve the large scale problems.

3. SOLUTION PROCEDURE

The particle swarm optimization (PSO), firstly introduced by Kennedy and Eberhart [8], is a population based optimizer that is inspired by the social behavior of a bird flock or fish school. Searching for food, every particle in the swarm adjusts its position according to the velocity which is determined by the particle velocity vector in the previous iteration, the best solution found by the particle as \(p_{best}\), and the best solution found by the whole swarm as \(g_{best}\). Thus each particle, as a potential solution for the optimization problem, memorizes and considers its own best past experience and the best experience gained by the other companions to make decision. Merit of any particle in the swarm is judged by its fitness value or objective function. As our model consists of various constraints, infeasibility of solutions is of high probability. To overcome this problem, a penalty value is assigned to infeasible particle so that its attributes will not be considered by others for future search. Improving the diversity of the swarm and preventing premature convergence, two uniform and non-uniform mutation operators are added to the algorithm. The pseudo-code is presented in Figure 3.

3.1. CHROMOSOME REPRESENTATION

We used the direct coding to represent the chromosomes. Each chromosome’s length is equal to \(i \times j \times t\), where \(i\) is the number of components, \(j\) is the number of suppliers, and \(t\) is the time periods. Every allele’s value is a real represents \(x_{ijt}\) and limited to \(C_{ij}\), the maximum capacity of the supplier \(j\), for component \(i\), in period \(t\). To avoid the complexity of the development, no other array is considered for \(Y\) as its values can be calculated according to \(X\), i.e. if \(\sum X_{ijt} > 0\) then \(Y_{ijt} = 1\) else \(Y_{ijt} = 0\).

3.2. PARTICLE VELOCITY AND POSITION
The particle movement is dependant to its velocity. The velocity leads the particle to the pbest and gbest solutions and prevents it from being trapped in local optima. The position and the velocity are calculated as follows:

\[ v_{ij}^t = w \times v_{ij}^{t-1} + c_1 \times r_1 \times (p_{best}^t_{ij} - X_{ij}^{t-1}) + c_2 \times r_2 \times (g_{best}^t_{ij} - X_{ij}^{t-1}) \]  

\[ X_{ij}^t = X_{ij}^{t-1} + V_{ij}^t \quad 1 \leq i \leq m, 1 \leq j \leq n, 1 \leq t \leq k \]

where \( m, n \) and \( k \) are the number of particles in the swarm, the dimension of the search space and the maximum number of iterations respectively. \( r_1 \) and \( r_2 \) are two random numbers in \([0, 1]\) which are used to effect the stochastic nature of the algorithm [9]. \( c_1 \) and \( c_2 \) are acceleration coefficients that lead the particle toward pbest and gbest and are commonly considered as constants in \([0, 2]\). \( w \) is the inertia weight and controls the effect of previous velocity on the current. Smaller value of \( w \) achieves the local exploitation while the global exploration is attained by larger values.

3.3. MUTATION OPERATORS

Mutation operator prevents the population from premature convergence to local optima. The developed PSO enjoys the uniform and non-uniform mutation operators for producing new diverse individuals. Also the probability of not being mutated is considered as well, i.e. individuals or particles are divided into three groups:

1. Those whose index divided by three are equal to zero, are mutated with non-uniform mutation.
2. Those whose index divided by three are equal to one, are mutated with uniform mutation.
3. Those whose index divided by three are equal to two, are not mutated.

3.3.1. Uniform Mutation

The uniform mutation is a basic operator in which all alleles have the probability to be mutated. For mutation, new value of the selected allele should be chosen from the lower and upper bound. In this paper, the new value is calculated as the following:

\[ y_i = x_i + (r - 0.5) \times \beta \]

where \( r \) is a random variable in \([0, 1]\) and \( \beta \) is a user defined parameter, which is set to 0.6 in this paper.

3.3.2. Non-uniform Mutation

There is a probability of disturbance for a good solution after a uniform mutation. Non-uniform mutation decrease the amount of mutating as the generations proceed. We have used the following function:

\[ y_i = x_i + (x_i^{u} - x_i^{l}) \times (1 - r_j) \times (1 - \frac{t}{t_{max}})^\beta \]

where \( x_i^{u} \) and \( x_i^{l} \) are the upper and the lower bound of \( x_i \), \( r_j \) is a random variable in \([0, 1]\), \( t_{max} \) is the maximum iteration number and \( \beta \) is a user defined parameter which determines the degree of dependency on iteration number (we used \( \beta = 0.6 \)).
4. NUMERICAL EXAMPLE

Consider a product with four components and three suppliers over a planning horizon of three periods. Each final product requires one unit of components 1 (Essential), 2 (Process) and 3 (Inventory), and two units of components 4 (Inventory). Transaction costs of three suppliers over a planning horizon of three periods are variable and are given in Table 1. Purchase costs of four components from three suppliers and the capacity of the suppliers for four components are shown in Table 2. Selling price \( S_t \), demands \( D_t \) and shortage cost \( SC_t \) of product are dynamic and are presented in Table 3. The model is solved using GAMS and PSO, and the results are presented in Table 4. As the problem is not large, the optimal solution is obtained by GAMS.

Table 1: Transaction costs of suppliers

<table>
<thead>
<tr>
<th>Suppliers</th>
<th>Planning horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 2: Purchase costs of components from three suppliers and suppliers’ capacities for four components over a planning horizon of three periods

<table>
<thead>
<tr>
<th>Components/Suppliers</th>
<th>Planning horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Purchase cost</td>
</tr>
<tr>
<td>1.1</td>
<td>60</td>
</tr>
<tr>
<td>1.2</td>
<td>70</td>
</tr>
<tr>
<td>1.3</td>
<td>80</td>
</tr>
<tr>
<td>2.1</td>
<td>65</td>
</tr>
<tr>
<td>2.2</td>
<td>55</td>
</tr>
<tr>
<td>2.3</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>30</td>
</tr>
<tr>
<td>4.1</td>
<td>25</td>
</tr>
<tr>
<td>4.2</td>
<td>20</td>
</tr>
<tr>
<td>4.3</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 3: Selling price, demands and shortage cost of product over a planning horizon of three periods

<table>
<thead>
<tr>
<th>Planning horizon</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_t )</td>
<td>2000</td>
<td>2100</td>
<td>2200</td>
</tr>
<tr>
<td>( D_t )</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>( SC_t )</td>
<td>400</td>
<td>500</td>
<td>600</td>
</tr>
</tbody>
</table>

It is important to note that for small sized problems the optimal solution can be achieved by GAMS whereas for large sized problems only a feasible solution might be found while meta-heuristic techniques like PSO are capable of finding near optimal solutions.

5. CONCLUSION

In this paper, the problem of an assembly system in a supply chain with multiple capacitated suppliers over multiple time periods was studied. Shortage of components in some periods brings about different scenarios which were formulated as a mixed integer mathematical model with regard to variability of prices, suppliers’ capacities, setup, holding and delay costs. The model is a combination of inventory lot-sizing and supplier selection and cannot
be solved in a reasonable computation time for large scale problems; therefore a PSO based algorithm was developed to solve the model. Finally a sample problem was presented and solved using GAMS and the implemented PSO. The proposed model can be extended to the multi product case, and different discipline can be defined for allocation of components to incomplete products as well.

Table 4: Numerical example results

<table>
<thead>
<tr>
<th>Variable</th>
<th>GAMS results</th>
<th>PSO results</th>
<th>Variable</th>
<th>GAMS results</th>
<th>PSO results</th>
<th>Variable</th>
<th>GAMS results</th>
<th>PSO results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{11}</td>
<td>1</td>
<td>1</td>
<td>X_{131}</td>
<td>2</td>
<td>2.822</td>
<td>X_{321}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y_{12}</td>
<td>1</td>
<td>1</td>
<td>X_{132}</td>
<td>0</td>
<td>0</td>
<td>X_{322}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y_{13}</td>
<td>0</td>
<td>0</td>
<td>X_{133}</td>
<td>3</td>
<td>3.2131</td>
<td>X_{323}</td>
<td>4</td>
<td>3.001</td>
</tr>
<tr>
<td>Y_{21}</td>
<td>0</td>
<td>0</td>
<td>X_{211}</td>
<td>0</td>
<td>0</td>
<td>X_{331}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y_{22}</td>
<td>0</td>
<td>0</td>
<td>X_{212}</td>
<td>5</td>
<td>6</td>
<td>X_{332}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y_{23}</td>
<td>1</td>
<td>1</td>
<td>X_{213}</td>
<td>0</td>
<td>0</td>
<td>X_{333}</td>
<td>4</td>
<td>4.999</td>
</tr>
<tr>
<td>Y_{31}</td>
<td>1</td>
<td>1</td>
<td>X_{221}</td>
<td>0</td>
<td>0</td>
<td>X_{411}</td>
<td>7</td>
<td>8.01</td>
</tr>
<tr>
<td>Y_{32}</td>
<td>0</td>
<td>0</td>
<td>X_{222}</td>
<td>0</td>
<td>0</td>
<td>X_{412}</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Y_{33}</td>
<td>1</td>
<td>1</td>
<td>X_{223}</td>
<td>3</td>
<td>3.0071</td>
<td>X_{413}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{111}</td>
<td>4</td>
<td>3.683</td>
<td>X_{231}</td>
<td>6</td>
<td>5.9002</td>
<td>X_{421}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{112}</td>
<td>5</td>
<td>6.7</td>
<td>X_{232}</td>
<td>0</td>
<td>0</td>
<td>X_{422}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{113}</td>
<td>0</td>
<td>0</td>
<td>X_{233}</td>
<td>5</td>
<td>4.993</td>
<td>X_{423}</td>
<td>12</td>
<td>13.07</td>
</tr>
<tr>
<td>X_{121}</td>
<td>0</td>
<td>0</td>
<td>X_{311}</td>
<td>5</td>
<td>5</td>
<td>X_{431}</td>
<td>9</td>
<td>7.99</td>
</tr>
<tr>
<td>X_{122}</td>
<td>0</td>
<td>0</td>
<td>X_{312}</td>
<td>6</td>
<td>6.002</td>
<td>X_{432}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{123}</td>
<td>5</td>
<td>4.787</td>
<td>X_{313}</td>
<td>0</td>
<td>0</td>
<td>X_{433}</td>
<td>4</td>
<td>3.93</td>
</tr>
</tbody>
</table>

Z = 25530.00 22433.70

REFERENCES

Developing a Mathematical Model to Evaluate and Select the Supplier/s in Presence of Multiple Price Discounts and Budget Limitation

Jafar Razmi¹, Elham Maghool
Department of Industrial Engineering, University College of Engineering, University of Tehran, P.O. Box: 11155-4563, Tehran, Iran,

ABSTRACT
Supplier selection is one of the essential issues in supply chain management. The importance of this issue is appeared because multiple criteria for selecting the best supplier/s must be considered. In the real world situation, each supplier may offer different discount schemes like all-unit discount or incremental discount. Furthermore, each supplier may offer a delay of payment in order to increase the number of buyer. Particularly, when there are not enough budgets for payment of purchased item, the delay of payment plays an important role in supplier selection problem. In this article, a single item, multi-supplier, multi-period purchasing problem with different quantity discount policy and various type of delay in payment is considered. An integer programming model with the object of minimizing the purchasing cost, under dynamic demand condition and budget limitation is formulated for this problem.

1.INTRODUCTION

Since the optimal decision in supplier selection problem is reduced the cost of purchasing material which is constituted the main cost of final product, the supplier selection problem is considered as one of the most important issue in supply chain management. The supplier selection problem is declared as a multi-objective model, when multiple criteria such as price, quality and delivery are considered simultaneously in this problem. In this circumstance, the tradeoff between tangible and intangible criteria to make an optimal decision is complicated. When a buyer is faced with a limited budget to purchase materials in each period, the delay of payment which is proposed by suppliers is considered as a best option to address this problem.

There are many ways to encourage buyers to procure large quantities of materials. One of the most influential ways is discount proposed by the suppliers. There are two types of price discount. One of these, is depend on the total quantity or variety of product purchased over a given period of time, another one is depend on the total value of sales volume (Xia and Wu, 2007). In Supplier Selection Problem (SSP), each supplier may follow new strategies to achieve more profit. Different kinds of payment for purchased items and various types of discount schedules (all-unit discount and/or incremental discount) may be introduced by each supplier in each period simultaneously. Hence, the selection of suppliers and the determination of order quantities and order points to be placed with the selected suppliers is a complex decision making problem.

In this paper, a mathematical model for a single item, multiple supplier, multi-period purchasing problem by considering two discount schemes and different types of payments simultaneously. The objective function of proposed model is minimizing the cost of purchasing item. The remainder of this paper will be organized as follows. In section 2, some previous studies and relevant literature are reviewed. In section 3, new developed model formulation of supplier selection problem will be presented and the result of numerical example will be shown in section 4. Finally conclusions and future researches will be outlined.

¹ Corresponding author: Tel: +9821-88021067; Fax: +9821-88013102
Email Addresses: jrazmi@ut.ac.ir (J. Razmi)
2. LITERATURE REVIEW

Dahel [4] considered multi-objective supplier selection problem in a multiple-product, multiple-supplier competitive sourcing environment. In this context, the vendors offer discounts on total amount of sales volumes. Price, delivery, and quality are considered as objectives in this problem. Torres and Mahmoodi [13] considered a decision model for allocating demand across a set of suppliers by considering supplier failure to satisfy the demand. In this model the purchasing cost and the cost of maintaining a set of suppliers and demand allocation are optimized. Also, the ability of other suppliers to satisfy the demand in the case of multiple sourcing is taken into account in this model. Lung Ng [8] applied a simple model for multi-criteria supplier selection problem which can be easily implemented with a spreadsheet package. Demirtas and Ustun [14], proposed an integrated approach of analytic network process (ANP) and multi-objective mixed integer linear programming to consider both tangible and intangible factors in choosing the best suppliers and best order quantities to maximize the total value of purchasing and minimize the budget and defect-rate.

Gossens et al. [6] studied the procurement problem by considering total quantity discount policy. A mathematical formulation is applied for this problem. They categorized this problem in four groups. In the first case, the market share of each supplier is considered. In the second case, the allowance of the buyer to procure more goods than strictly needed is studied. In the third case, the number of winning supplier is limited, and in the last case, a multi-period model with inventory cost is considered. All of these problems can be solved by defining a series of min-cost flow problems.

Susuz and Kokangul [11] considered an integration of analytical hierarchy process and non-linear integer and multi-objective programming under some constraint such as quantity discounts, capacity and budget. Both qualitative and quantitative factors are considered in this supplier selection problem. The objectives of this model are maximizing the total value of purchase (TVP), minimizing the total cost of purchase (TCP) or maximizing TVP and minimizing TCP. Burke et al. [2], considered the optimal sourcing policy for a single buyer under a variety of supplier pricing schemes including linear discounts, incremental unit discounts, and all unit discounts and supplier capacity limitations. In this paper, a heuristic solution methodology is developed to identify a quantity allocation decision for the buyer.

Ghodsypour and O'Brien [5] proposed a mixed integer programming model to solve multiple sourcing problems, by considering the total cost of logistics, including net price, storage, and transportation and ordering costs. They take into accounts buyer limitations on budget, quality and services. Wadhwa and Ravindran [10] considered the vendor selection problem as a multi-objective optimization problem, where each buyer order multiple products from different vendors in a multiple sourcing network. Three conflicting criteria such as price, lead –time and rejects are defined for the problem. A pricing model under quantity discounts is used to represent the purchasing cost. Several methods such as weighted objective goal programming, and compromise programming are applied for solving this problem.

Tempelmeier [12], proposed a simple heuristic for a supplier selection and purchase order sizing for a single item problem in dynamic demand environment. In the problem, each supplier can offer all-unit or incremental quantity discounts which may vary overtime. The object of this model is minimizing the ordering cost of purchasing item and holding cost. Basnet et al. [1], considered a multiple-period inventory lot-sizing with supplier selection problem where there are multiple products and multiple supplier. This problem is formulated by a mixed integer programming model. The objective function consist of purchasing cost of products, the transaction cost of suppliers and holding cost for remaining inventory in each period. An enumerative search algorithm and two heuristics are applied for addressing this problem. Chang [3], addressed the single-item, multi-supplier, multi-period problem with real world constraint by proposing an exact acquisition policy. The object of this problem is minimizing the periodic purchasing cost, ordering cost, and holding cost. Also, a variety of price quantity discount policies which is proposed by suppliers is considered in the paper. Mohammad Ebrahim and Razmi [9] considered a mathematical model for supplier selection and single item purchasing problem by considering different discount schemes (all-unit discount, incremental discount, and total business volume discount) simultaneously. They proposed a scatter search algorithm to solve this problem.

It is clear that none of the literature has focus on analysis of the cost when buyer/s is faced with different discount schedule proposed by supplier/s. In addition, practitioners should be helped in the case of offering variety of payment schedule. In this paper, the single-item, and multi-period supplier selection problem under limited budget condition is considered. In real world situation, each supplier may offer a special discount policy and proposes one type of payment which is defined in previous section. A mathematical model with the object of minimizing the cost function is proposed for this problem. The result of solving example shows that the delay of payment is very important criteria in supplier selection problem particularly when the buyer faces with limited
budget in each period. Although each supplier may offers a good discount schedules in each period, but the delay of payment plays an important role in supplier selection decision. By considering both quantity discount schedules which are offered by supplier and different type of delay in payment which is applied in supplier selection problem, the effectiveness of each criterion is tested and the importance of each criterion is conducted.

3. PROPOSED MODEL

In this paper, a new model formulation for the problem of supplier selection and order sizing under dynamic demand condition by considering different types of discount policies and delay of payments and the object of minimizing the purchasing cost subject to the following assumption:

1. Dynamic deterministic demands.
2. Only one item is purchased from suppliers.
3. The demand and purchased quantities in each period are integer.
4. The buyer is faced with limited budget in purchasing item.
5. There are several suppliers, each one may offer
   a. Time varying deterministic prices and time varying quantity discount schedules (all-unit and/or incremental discounts).
   b. Supplier -specific delivery periods.
   c. Supplier- specific delivery lead -times.
   d. Each supplier may offer different types of payment. Three kinds of payment are considered in this model.
      • The total amount of purchased item is paid at the order point.
      • The percentage of the total amount of purchased item is paid at the order point and the remaining amount is paid at the delivery point.
      • The percentage of the total amount of purchased item is paid at the delivery point and the remaining amount is paid at the next months after delivery point. (This delay of payment is determined with each supplier in each period.)
   e. Supplier –specific delay of payment.
   f. The rate of interest which is determined in each period and related with each supplier.
   g. No capacity constraint.

4. VARIABLES AND EQUATIONS

\( i \) \quad \text{index of suppliers}

\( k \) \quad \text{index of discount intervals}

\( x_{ik} \) \quad \text{purchased quantity from supplier} \ i \ \text{in discount interval} \ k \ (\text{integer variable})

\( y_{ik} \) \quad \text{binary variable; if the purchased quantity from supplier} \ i \ \text{falls on the interval} \ \text{Corresponding to this variable then} \ y_{ik} = 1, \ \text{otherwise} \ y_{ik} = 0

\( n \) \quad \text{total number of suppliers}

\( m_{ij} \) \quad \text{Number of suppliers which offer all unit discounts and followed the first method of payment.}

\( m_{i2} \) \quad \text{Number of suppliers which offer all unit discounts and followed the second method of payment.}

\( n_1 \) \quad \text{Number of suppliers that offer all-unit discount and followed the third method of payment.}

\( m_{i2} \) \quad \text{Number of suppliers which offer incremental discount and followed the first method of payment.}

\( m_{i2} \) \quad \text{Number of suppliers which offer incremental discount and followed the second method of payment.}

\( n_2 \) \quad \text{Number of suppliers that offer incremental discount and followed the third method of payment.}

\( l_{ik} \) \quad \text{lower bound of the discount interval} \ k \ \text{offered by supplier} \ i

\( u_{ik} \) \quad \text{upper bound of the discount interval} \ k \ \text{offered by supplier} \ i

\( p_{ik} \) \quad \text{discounted unit price of the discount interval} \ k \ \text{offered by supplier} \ i

\( k_i \) \quad \text{index of the last interval offered by supplier} \ i

\( d_t \) \quad \text{Net requirements for the item in period} \ t

\( l_i \) \quad \text{the lead time of item which is supplied by supplier} \ i

\( \theta_i \) \quad \text{the delay of payment which is declared by supplier} \ i
\[ i_\tau \] the interest rate in period \( \tau \)
\[ B \] the total budget of buyer for all period
\[ \delta_{ik}^\tau \] Proportion of demand in period \( t \), that is delivered by supplier \( i \) in period \( \tau \) with Discount level \( k \)
\[ r_i^\tau \] The percentage of payment which is paid to supplier \( i \) in period \( \tau \).

\[
\begin{align*}
\text{max} & \quad \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau + \sum_{\tau = 1}^{T} \sum_{i = m_{i+1}}^{m_i} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau (1 - r_i^\tau)(1 + i_\tau)^{l_i} \\
& \quad + \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} p_{ik}^\tau r_i^\tau (1 + i_\tau)^{l_i} \\
& \quad + \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau (1 - r_i^\tau)(1 + i_\tau)^{\theta + l_i} \\
& \quad - \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau))] \\
& \quad + \sum_{\tau = 1}^{T} \sum_{i = m_{i+1}}^{m_i} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau)](1 - r_i^\tau)(1 + i_\tau)^{l_i} \\
& \quad + \sum_{\tau = 1}^{T} \sum_{i = m_{i+1}}^{m_i} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau)](1 + i_\tau)^{l_i} \\
& \quad + \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau)](1 - r_i^\tau)(1 + i_\tau)^{\theta + l_i}
\end{align*}
\]

Subject to:

Budget limitation:

\[
\sum_{\tau = 1}^{T} \sum_{i = 1}^{m_1} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau + \sum_{\tau = 1}^{T} \sum_{i = m_{i+1}}^{m_i} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau + \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} p_{ik}^\tau x_{ik}^\tau (1 - r_i^\tau)(1 + i_\tau)^{l_i} \\
+ \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau)](1 - r_i^\tau)(1 + i_\tau)^{l_i} \\
+ \sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} [p_{ik}^\tau (x_{ik}^\tau - y_{ik}^\tau u_{ik,i-1}^\tau) + y_{ik}^\tau \sum_{j = 1}^{k-1} p_{ij}^\tau (u_{ij}^\tau - u_{ij-1}^\tau)](1 + i_\tau)^{l_i} \leq B
\]

Demand fulfillment:

\[
\sum_{\tau = 1}^{T} \sum_{i = 1}^{n} \sum_{k = 1}^{k_f} \delta_{ik}^\tau = 1 \quad t = 1, \ldots, T
\]
Definition of order sizes:

\[ \sum_{\tau=1}^{T} \delta_{ik}^{\tau} d_{\tau} = x_{ik}^{\tau} \quad \forall i = 1,\ldots,n \quad \tau = 1,\ldots,T \quad k = 1,\ldots,k_{i}^{\tau} \]  

(4)

Definition of discount levels for the selected delivery period which is declared the lower and upper bound of discount levels.

\[ \delta_{ik}^{\tau} \geq 0 \quad \forall \tau = 1,\ldots,T \quad t = \tau, \tau + 1,\ldots,T \quad i = 1,\ldots,n \quad k = 1,\ldots,k_{i}^{\tau} \]  

(5)

\[ \delta_{ik}^{\tau} \leq y_{ik}^{\tau} \quad \forall i = 1,\ldots,n \quad \tau = 1,\ldots,T \quad t = \tau, \tau + 1,\ldots,T \quad k = 1,\ldots,k_{i}^{\tau} \]  

(6)

\[ l_{ik}^{\tau} y_{ik}^{\tau} \leq x_{ik}^{\tau} \leq u_{ik}^{\tau} y_{ik}^{\tau} \quad \forall i = 1,\ldots,n \quad k = 1,\ldots,k_{i}^{\tau} \quad \tau = 1,\ldots,T \]  

(7)

\[ \sum_{k=1}^{k_{i}^{\tau}} y_{ik}^{\tau} \leq a_{i}^{\tau} \quad \forall i = 1,\ldots,n \quad \tau = 1,\ldots,T \]  

(8)

\[ a_{i}^{\tau} = \begin{cases} 1 & \text{if supplier I can deliver in period} \\ 0 & \text{else} \end{cases} \]  

(9)

\[ x_{ik}^{\tau} \in \mathbb{Z}^{+} \cup \{0\} \quad \forall i = 1,\ldots,n \quad k = 1,\ldots,k_{i}^{\tau} \quad \tau = 1,\ldots,T \]  

(10)

\[ y_{ik}^{\tau} \in \{0,1\} \quad \forall i = 1,\ldots,n \quad k = 1,\ldots,k_{i}^{\tau} \quad \tau = 1,\ldots,T \]  

(11)

5. NUMERICAL EXAMPLE

By considering all cases which may be exist in real world situation, these examples are defined. In the first one, suppose there are three suppliers in which each supplier consider one kind of payment. One of them offers incremental discount and other suppliers offer all-unit discount. In this example, the price structure which is considered for competitors is similar and no supplier dominates the others. Suppose that the offered discount schedules are similar in all periods. The demand series were generated based on a gamma-distribution with different coefficient of variation between 0.2 and 4.05 with the mean values between 37 and 64. In this example, six suppliers are considered, in which supplier A1, select the first type of payment and offers all-unit discount policy. Supplier A2 offers all-unit discount schedule and select the second type of payment. Supplier A3 offers all-unit discount scheme and consider the third type of payment. Supplier A4 selects the first type of payment and offers incremental discount. Supplier A5 offers incremental discount and choose the second type of payment. Supplier A6 offers incremental discount and propose the third type of payment. The prices are offered in five levels with each supplier which are shown in Table 1.

The demands of item in each period which are generated by Gamma distribution are shown in Table 2. In this circumstance, by considering the limited budget supplier A3 and supplier A6 is selected by buyer to purchase a large amount of material for satisfying the demand. The selected suppliers offer a delay in payment which may be the best option to select by buyer. The result of purchasing quantities is shown in table3 and table6. The purchased quantity from supplier i in discount level j and in load period k is shown by \( x(i,j,k) \) in Table 3.
Table 1. Discount levels which are offered by each supplier.

<table>
<thead>
<tr>
<th>r</th>
<th>$l_k^1$</th>
<th>$p_{ik}^1$</th>
<th>$l_k^2$</th>
<th>$p_{ik}^2$</th>
<th>$l_k^3$</th>
<th>$p_{ik}^3$</th>
<th>$l_k^4$</th>
<th>$p_{ik}^4$</th>
<th>$l_k^5$</th>
<th>$p_{ik}^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>4.10</td>
<td>50</td>
<td>4.22</td>
<td>50</td>
<td>4.26</td>
<td>50</td>
<td>4.10</td>
<td>50</td>
<td>4.22</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
<td>3.72</td>
<td>125</td>
<td>3.79</td>
<td>125</td>
<td>3.76</td>
<td>125</td>
<td>3.72</td>
<td>150</td>
<td>3.79</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>3.55</td>
<td>250</td>
<td>3.48</td>
<td>250</td>
<td>3.51</td>
<td>250</td>
<td>3.55</td>
<td>300</td>
<td>3.48</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
<td>3.39</td>
<td>500</td>
<td>3.36</td>
<td>500</td>
<td>3.32</td>
<td>500</td>
<td>3.39</td>
<td>500</td>
<td>3.36</td>
</tr>
<tr>
<td>5</td>
<td>$\infty$</td>
<td>3.16</td>
<td>$\infty$</td>
<td>3.13</td>
<td>$\infty$</td>
<td>3.13</td>
<td>$\infty$</td>
<td>3.16</td>
<td>$\infty$</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Table 2. Demand for item in each period.

<table>
<thead>
<tr>
<th>period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>demand</td>
<td>45</td>
<td>34</td>
<td>50</td>
<td>38</td>
<td>62</td>
<td>62</td>
<td>49</td>
<td>52</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>period</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>demand</td>
<td>47</td>
<td>57</td>
<td>43</td>
<td>75</td>
<td>48</td>
<td>50</td>
<td>61</td>
<td>50</td>
<td>48</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 3. Purchased quantities from each supplier.

<table>
<thead>
<tr>
<th>variable</th>
<th>Purchased quantity</th>
<th>variable</th>
<th>Purchased quantity</th>
<th>variable</th>
<th>Purchased quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(3, 2, 1)</td>
<td>40.00000</td>
<td>X(3, 2, 7)</td>
<td>1.00000</td>
<td>X(5, 2, 1)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(3, 2, 2)</td>
<td>2.00000</td>
<td>X(3, 2, 8)</td>
<td>1.00000</td>
<td>X(5, 2, 2)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(3, 2, 3)</td>
<td>1.00000</td>
<td>X(3, 2, 9)</td>
<td>1.00000</td>
<td>X(5, 2, 3)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(3, 2, 4)</td>
<td>1.00000</td>
<td>X(3, 2, 10)</td>
<td>1.00000</td>
<td>X(5, 2, 4)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(3, 2, 5)</td>
<td>40.00000</td>
<td>X(3, 2, 12)</td>
<td>1.00000</td>
<td>X(5, 2, 5)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(3, 2, 6)</td>
<td>1.00000</td>
<td>X(3, 2, 13)</td>
<td>1.00000</td>
<td>X(5, 2, 6)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(5, 2, 7)</td>
<td>1.00000</td>
<td>X(5, 2, 13)</td>
<td>1.00000</td>
<td>X(5, 2, 19)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(5, 2, 8)</td>
<td>1.00000</td>
<td>X(5, 2, 14)</td>
<td>1.00000</td>
<td>X(5, 2, 20)</td>
<td>1.00000</td>
</tr>
<tr>
<td>X(5, 2, 9)</td>
<td>1.00000</td>
<td>X(5, 2, 15)</td>
<td>1.00000</td>
<td>X(6, 5, 1)</td>
<td>301.0000</td>
</tr>
<tr>
<td>X(5, 2, 10)</td>
<td>1.00000</td>
<td>X(5, 2, 16)</td>
<td>1.00000</td>
<td>X(6, 5, 2)</td>
<td>301.0000</td>
</tr>
<tr>
<td>X(5, 2, 11)</td>
<td>1.00000</td>
<td>X(5, 2, 17)</td>
<td>1.00000</td>
<td>X(6, 5, 14)</td>
<td>301.0000</td>
</tr>
<tr>
<td>X(5, 2, 12)</td>
<td>1.00000</td>
<td>X(5, 2, 18)</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Objective function value: 465794.7
6. SUMMARY

In this paper, the single-item, and multi-period supplier selection problem under limited budget condition is considered. In real world situation, each supplier may offer a special discount policy and proposes one type of payment which is defined in previous section. A mathematical model with the object of minimizing the cost function is proposed for this problem. The result of solving two examples shows that the delay of payment is very important criteria in supplier selection problem particularly when the buyer faces with limited budget in each period. Although each supplier may offers a good discount schedules in each period, but the delay of payment play an important role in supplier selection decision. As shown in the example, three scenarios for delay in payment are considered. The buyer selects the two suppliers, A3 and A6 which offered a delay in payment. It shows that the third scenario is the best one for selection and the buyer may acquire more profit than selecting other scenarios. Therefore, when the buyer faced with the limited budget to satisfy the demand in spot market, he prefers to purchase the whole amount of materials from suppliers which proposed a proper delay of payment. By considering both quantity discount schedules which are offered by supplier and different type of delay in payment which is applied in supplier selection problem, the effectiveness of each criterion is tested and the importance of each criterion is conducted.

REFERENCES

An Investigation of Manufacturing Challenges in Automotive Software Development from a Lean Perspective

Joakim Pernstål and Peter Öhman

1 Computer Science and Engineering
Chalmers
SE-421 96 Gothenburg

ABSTRACT

The automotive industry is facing a tremendous growth in software and system engineering in product development. This is pressing the automakers to build up software competencies and elaborate efficient processes for product development and production. The involvement of manufacturing engineering in product development has consequently become imperative since the increase in software-based systems implies a greater number of product variants and a higher degree of complexity to be managed in the manufacturing processes.

This paper presents an investigation of the interaction between the functions of product development and manufacturing in automotive software development from a lean production perspective. The primarily objective of the study presented in this paper is to increase the understanding of the area by using complementary perspectives found in related literature and the empirical results of a previous explorative multiple case study undertaken at two Swedish automotive companies. Since automakers aim to achieve product differentiation in a broad market and cost leadership by adopting lean production emerged issues were primarily analyzed from a lean perspective. Challenging issues were found and the analysis indicated that it is useful to adopt a lean approach within the realm of the investigated area of interest since it seems to constitute an appropriate foundation for remedial measures.

1. INTRODUCTION

Trends in the automotive industry show that the number of electrical functions in vehicles is increasing and that they are becoming more complex as described in [1] and [2]. Current premium vehicles contain not less than 70 Electronic Control Units (ECU) with the requisite software that enables and controls various features and electrical systems. Moreover, [1] estimates that 80 percent of all future automotive innovations will be driven by electronics and 90 percent thereof by software and according to [2] it is expected that the upper class vehicles will have up to 1 GB of software in five years time. The tremendous increase in automotive electronics is for example driven by legislation and customers' demand for new functionalities such as reduced gas consumption and increased performance and safety. Moreover, software-enabled features facilitate product differentiation and customization which is an important competitive capability among automakers.

This evolution is pressing the automotive industry to elaborate efficient processes, obtain the necessary knowledge and skills among employees, and adopt appropriate tools and technologies supporting the development of software-based systems. The main goal of these activities is to achieve a quality assured and cost efficient launch of the vehicles in the manufacturing process when production starts. Automotive manufacturing processes primarily affected by software in vehicles can be divided into assembly plant software download (SWDL), where software-enabled features in the vehicle are configured by downloading application software or calibration data, and verification by electrical testing which secures that the vehicles are correctly assembled.

Meeting customers' demands for a broader range of products with various features and tighter scheduling of new model launches involve an increased number of products and systems that must be developed in a shorter time. The involvement of manufacturing engineering in product development has consequently become imperative since the increase in software-based systems implies a greater number of product variants and a higher degree of complexity to be managed in the manufacturing processes. Parallel to this, there is an ever increasing demand concerning vehicle
manufacturing capabilities such as cost, quality, and flexibility and delivery performance. Thus, to achieve a successful launch, the product and prerequisites for the manufacturing processes, e.g. line speed, tools, competence and man power, must be harmonized, and this requires a well functioning interaction between the functions of manufacturing and product development (PD)

The point of departure of the study presented here is an explorative multiple case study by [3] that was undertaken at two Swedish automotive companies and focused on the interaction between the functions of manufacturing and PD in the development of software-based systems in the automotive domain. In this paper the manufacturing challenges in the area of interest (issues) that emerged in the analysis of the data collected was supplemented with a review of pertinent literature and an extended analysis using complementary perspectives found in related literature. To obtain a top-down view, appropriate overall theories related to the research area were reviewed and decomposed to find links between inherent concepts, principles and practices, and the issues identified in the case studies conducted. The review acknowledges that an overall view among researchers, e.g.[4] and [5], is that a goal in the automotive industry is to simultaneously achieve product differentiation on a broad market and cost leadership by adopting a lean approach to manufacturing and PD. Consequently, the analysis of the issues identified in the field studies conducted is primarily based on the lean concept and inherent principles resulting in recommendations accompanying the identified issues for adequate implementation of improvements.

The paper is organized as follows. Section 2 provides a brief literature review covering pertinent literature related to lean concept and principles. Section 3 analyses and discusses the issues identified from a lean perspective resulting in a framework composed of recommendations for remedial measures. Finally, conclusions are presented in section 4.

2. LITERATURE REVIEW

Outcomes of studies of the automobile industry by researchers at the International Motor Vehicle Program (IMVP) at the Massachusetts Institute of Technology (MIT) showed that the lean approach developed at and adopted by Japanese automakers outperformed that of Western companies considering quality, productivity and flexibility [6]. In particular, the Toyota Production System (TPS) constitutes one example in the automotive industry where Japanese production concepts have been elaborated to become the most competitive production systems in the automotive industry. The story of TPS has been described for example by [7] as a production system that has gradually and cumulatively evolved since the 1940s and is essentially a hybrid containing elements from the Ford system that have been integrated in a domestic environment. The system initially focused on production activities. However, the principles of TPS are also applicable to PD, customer management and supplier management and logistics.

The most popular publication considering Lean Production (LP) is the book The Machine that Changed the World [4] which depicts in its enterprise model of lean competitors a lean approach that includes the entire organization and not just the manufacturing units. The model provides a comprehensive view of the management system at Toyota and emphasizes that the power of the LP concept depends on the ability to combine the five elements of a lean business system, which are: 1) the product development process, 2) the supplier management process, 3) the manufacturing management process, 4) the customer management process and 5) managing the combined enterprise. Based on more than 20 years of studies of TPS "The Toyota Way" [5] also points out that LP comprises more than TPS. Liker [5] emphasizes that LP is more a philosophy throughout the whole enterprise than an implementation of lean methods and tools where lean producers stress innovativeness, high reliance on relatively autonomous teams of employees that possess broad training, experience and learning, and continuous improvement as requisites for world class automotive manufacturing. Further, capability building in lean producing firms is augmented by incremental bottom-up decision-making processes in a stable environment. This philosophic view however implies that lean production may be suitable in Japanese industry owing to its historical and cultural heritage, which is acknowledged by [8]. Although it took decades for Toyota and other Japanese auto manufacturers to develop and implement the LP concept, Western companies try to accomplish the same task in a few years or even less. Nevertheless, in meeting the fierce competition, it is a common view among researchers and the automotive industry that it is necessary to develop and adopt a lean thinking in these organizations.

Womack et al. [4] introduced the Lean Product Development (LPD) concept and concluded that there were differences between mass and Japanese lean producers in the areas of leadership, teamwork, communication and simultaneous development. The principles for LPD are based on the Toyota Product Development System (TPDS) originating from the heavyweight management system which is one of the core elements of TPS. Toyota adopted the system in the 1950s and called it the shusa system where the shusa is a strong product manager acting as concept
creator and project coordinator at the same time. Womack et al. [4] compare the shusa or Chief Engineer (CE) to a new super craftsman.

Morgan and Liker [9] suggest that LPD involves more than lean methods. Using Sociotechnical Systems Theory (STS) and the principles of TPDS, [9] developed a Lean Product Development System (LPDS), which is presented in Figure 1. STS theorists such as [10] emphasize that work involves a combination of social and technical requisites that must be considered equally important. In contrast, rational system theorists such as [11] primarily view human workers as necessarily being adapted to the requirements of the technical solutions. Further, the STS theory comprises an open system approach. Many theories have been developed on the basis of this perspective. For example, the contingency theory developed by [12] is guided by the general hypothesis that organizations whose internal features match the demands of their environment achieve the best adaptations. In other words, an open system approach pays attention to interdependencies between actors in organizations and the elements outside those organizations' boundaries. The LPDS model contains three primary subsystems, 1) process, 2) people and 3) tools and technology, which are described by means of 13 principles.

Further, [9] define four categories of new product development (NPD) ranging from invention with totally new technology to upgrades of existing products: 1) new products that represent radically different products or technology, 2) revolutionary product platform development projects that require fundamentally new systems and components, 3) derivative products built on existing product platforms and 4) incremental product improvement. The LPDS model is primarily based on the third category since this is the most common approach to new programs in lean automotive companies.

3. ANALYSIS AND DISCUSSION OF ISSUES FROM LEAN PRODUCT DEVELOPMENT PERSPECTIVE

This section provides analyses and discussions of possible links between the issues identified in the multiple case study presented in [3] and lean principles and practices in PD. The primary data source used for the case studies were 20 in-depth interviews with stakeholders involved in the interaction between PD and manufacturing in the development of software-based systems since this qualitative method, according to Robson [13], has the potential of providing rich and highly illuminating material. In addition to the interviews, pertinent documentation and archive records were used to augment and corroborate interview data and triangulate data sources. The LPDS model developed by [9] was found appropriate as a foundation for the analysis from a LPD perspective of the issues identified in the studied organizations. Further, in order to specifically address the research area, the analysis focuses mainly on the primary issues found in the multiple case study. Although the relationships between these issues and LPDS principles were found to be somewhat intertwined, Table 1 shows the linkages that were conceived as most coherent, leading to a set of recommendations for further development of improvements. In-depth discussions and analyses of the links between every single issue and principles leading to the recommendations shown in Table 1 are not included in this paper for space limitations. To cover the three subsystems in the LPDS model, primary issues with linkages to principles representing the subsystems were selected and will be explicitly discussed in the following.

Issue 2 Lack of processes that secure design for vehicle configuration and testability in production: Participants at both companies brought up this issue and, to be able to introduce remedies that maximize overall performance in PD and manufacturing, it is necessary to gain a comprehensive understanding of the driving forces fostering the development of PD and manufacturing systems. A literature survey of PD (see for example [14])
indicates that customers’ demands for a broader range of products with various functions and tighter scheduling of new model launches involve an increased number of products and systems that must be developed in a shorter time.

Table 1: Linkages between empirical results, LPDS principles and recommendations

<table>
<thead>
<tr>
<th>Issue no</th>
<th>Description</th>
<th>Referred principles in the LPDS model</th>
<th>Recommendations based on principles in LPDS model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>There are difficulties in incorporating production requirements in the development of software-based systems</td>
<td>Principle 4: Utilize rigorous standardization to reduce variation, and create flexibility and predictable outcomes Principle 11: Adapt technology to fit your people and processes</td>
<td>* Establish checklists that are continuously updated with current prerequisites for production * Consider processes and people in the development of RE tools</td>
</tr>
<tr>
<td>2</td>
<td>Lack of processes that secure design for vehicle configuration and testability in production</td>
<td>Principle 2: Front-Load the Product Development Process while there is maximum design space to explore alternative thoroughly</td>
<td>* Consider the lean SBCE concept when developing processes and practices</td>
</tr>
<tr>
<td>3</td>
<td>Prerequisites for appropriate verification of manufacturing processes that are affected by SW in vehicles are not fulfilled</td>
<td>Principle 7: Develop towering technical competence in all engineers Principle 11: Adapt technology to fit your people and processes</td>
<td>* Develop tools that makes the software-based functions more tangible in prototyping e.g. model based development * Participation from manufacturing is important when tools for prototyping software-based functions are developed * Consider people and existing processes in the development of tools for model-based development</td>
</tr>
<tr>
<td>4</td>
<td>Different manufacturing processes and tools in the plants aggravate the adaptation between product and production Processes and instructions for quality assured release of software are not always adopted</td>
<td>Principle 4: Utilize rigorous standardization to reduce variation, and create flexibility and predictable outcomes</td>
<td>* Standardize manufacturing processes and tools since it is the foundation for continuous improvement (Kaizen) * Encourage process discipline since stability and standardization are the foundation for continuous improvement and flexibility</td>
</tr>
<tr>
<td>5</td>
<td>Manufacturing processes and tools that are affected by SW in vehicles do not fulfill desired efficiency and reliability</td>
<td>Principle 9: Build in learning and continuous improvement Principle 13: Use powerful tools for standardization and organizational learning</td>
<td>* Facilitate diffusion of knowledge leveraging continuous improvements (Kaizen) by a bottom-up approach * Develop adequate and powerful tools and models facilitating optimized adaptation of products and manufacturing processes</td>
</tr>
<tr>
<td>6</td>
<td>Lack of SW competence in the manufacturing organization Too little knowledge about the manufacturing processes in the PD organization leads to difficulties in introducing new products in production</td>
<td>Principle 9: Build in learning and continuous improvement Principle 7: Develop towering technical competence in all engineers</td>
<td>* Level out technical capabilities between design and manufacturing engineers by equalizing the conditions for recruitment and development * Consider manufacturing capabilities when recruiting and developing design engineers * Consider the ability to create learning networks in the formal organization</td>
</tr>
<tr>
<td>7</td>
<td>Lack of SW competence in the manufacturing organization Too little knowledge about the manufacturing processes in the PD organization leads to difficulties in introducing new products in production</td>
<td>Principle 9: Build in learning and continuous improvement Principle 7: Develop towering technical competence in all engineers</td>
<td>* Level out technical capabilities between design and manufacturing engineers by equalizing the conditions for recruitment and development * Consider manufacturing capabilities when recruiting and developing design engineers * Consider the ability to create learning networks in the formal organization</td>
</tr>
<tr>
<td>8</td>
<td>There are problems in mapping the manufacturing organization to the PD organization that lead to difficulties in obtaining multidisciplinary project teams</td>
<td>Principle 6: Organize to balance functional expertise and cross-functional integration</td>
<td>* Consider the ability of the formal organization to build informal networks in the mapping of functional organizations and product-oriented organizations</td>
</tr>
</tbody>
</table>

Thus, companies’ ability to reduce time-to-market has become one of the most important of their capabilities. To accelerate the PD process, a number of authors in the literature, for example [14], propose a more concurrent and multi-disciplinary approach to the development of products that maximize the use of enabling technologies such as rapid prototyping and cross-functional techniques, among them quality function deployment (QFD), failure mode and effect analyses (FMEA), Design for Manufacture and Assembly (DFMA) etc. Hence, to be competitive, the development of new products in the old sequential fashion “over the wall approach” has become insufficient.
A concurrent engineering approach is one of the cornerstones of the LPDS model (principle 2). In a study by [15] it was discovered that Toyota considers sets of possible designs in the early development phases, which gradually converge to one final design. They termed this approach Set-Based Concurrent Engineering (SBCE). In contrast to SBCE, they observed that U.S automakers used a refined version of traditional serial engineering. This approach is termed Point-Based Concurrent Engineering [15]. Condensed from [15], Figure 2 depicts the characteristics of SBCE with its parts A to D simplified by only two functions: design engineering and manufacturing engineering.

![Figure 2: SBCE condensed from [15]](image)

In contrast to the point-based iterative approach, SBCE is a gradual converging process. It starts in part A, where expertise from design engineering and manufacturing engineering define their broad sets of possible solutions. The convergence process carries on to part D, where the two functions continue to work in a concurrent manner by refining and narrowing the intersecting set of alternatives to a superior solution. SBCE involves more than parallel development of alternative solutions since it: 1) emphasizes rapid rejection of weak designs by ranking alternative systems in trade-off curves and decision matrices and 2) creates knowledge bases defining design spaces, i.e. limits of design possibilities.

In comparison to the iterative and the traditional serial approach, it is argued in [15] and [9] (principle 2) that SBCE reduces the risk of rework in late development phases and thus gives more robust and optimized systems. Likewise, in software development, [16] suggests that a set-based approach is preferable since it is an efficient way of communicating constraints for core concepts, leading to faster and better solutions and an amplification of the learning process in development teams. Certainly, it can be discussed whether the above described approach used in LPD is the most suitable for the studied environment. Nevertheless, an SBCE approach may serve as a recommendation for improvements of the activities in PD as concerns design for vehicle configuration and testability of software-based systems in vehicle production.

**Issue 3 Prerequisites for appropriate verification of manufacturing processes that are affected by SW in vehicles are not fulfilled:** One of the major phases in LPD incorporates prototype building and tooling with lean manufacturing. This phase aims to eliminate later engineering changes, i.e. "fixes", since it is the last stage in the PD process allowing any rework of products, and manufacturing processes as well. The prototype and tooling phase at lean producers is characterized by intensive cross-functional interaction in established core teams based on analyses of both physical and virtual prototypes. The core team includes different functions, such as manufacturing engineers, team leaders, tooling and designers, that have daily discussions about the progress of the build where all members are expected to answer quickly. In practice, the influence of software-based functions on the manufacturing processes is primarily interpreted and verified in trial production on physical build vehicles in the late stages of PD implying high cost and a limited number of iterations. However, as brought up by the respondents, the conditions for proper verification of associated manufacturing processes are not always fulfilled. This leads often to late changes ("fixes") of products and manufacturing processes that impair the manufacturing performance at the worst possible time, when the market demand is peaking after the launch of a new product.

Morgan and Liker [9] highlight the importance of the prototype and tooling phase in LPDS principle 7. The work in this stage of the PD process is particularly influenced and guided by the lean practice genchi gebutsu engineering which constitutes one of the four core principles in "The Toyota Way" [5]. Genchi gebutsu is translated in English as "Go and see for yourself to thoroughly understand the situation". The essence of this lean engineering practice is that engineers responsible for a specific situation are required to observe the real scene in which it is going on and analyzing it by revealing possible problems and their causes based on their knowledge and experience. Hence, to elaborate information facilitating discussions and appropriate decision-making in the core teams, it is not sufficient
to take anything for granted or to only rely on other engineers’ analyses and reports. Likewise, in software development, [16] emphasizes regular prototyping since it creates a need for cross-functional communication and provides early feedback.

Obtaining efficient prototyping enabling the above lean practice in the verification process of software-based systems and associated manufacturing processes is a challenging task, however, owing to two broad factors. First, in contrast to mechanical parts, which in early stages of the development can be visualized in applications such as CAD/CATIA or as physical prototypes, software-based systems are intangible and are often described in written specifications implying a higher abstraction level to be interpreted and understood. In addition, the software-based systems enable an increase of variants. According to [2], a premium car typically has about 80 electronic fittings that can be ordered depending on the country etc. Simple yes/no decisions for each function yield a possible maximum of roughly $2^{80}$ variants to be ordered and produced for a car. The second factor considers the manufacturing processes, tools and applications that are affected by software embedded in vehicles. The characteristics of these systems are often recognized as complex since they incorporate for example, vehicle communication technologies, processes including operators, interaction with other IT systems and wireless data transfer. Moreover, the systems have to fulfill production demands such as user friendliness, efficiency, availability, reliability and maintainability. Being able to understand and foresee the impact on the production of products and systems with high complexity described in abstract form and complex manufacturing systems also requires deep knowledge and a great deal of experience among both manufacturing engineers and developers.

To accomplish earlier prototyping that would facilitate earlier process verification and shorter feedback loops at a reduced cost, the participants at both companies suggested integration of the manufacturing processes in model-based development. At present, there a number of approaches and techniques used to modeling software-based systems in the automotive domain. For example, Unified Modeling Language (UML) [17], an objective-oriented approach is used in abstract modeling aiming to demonstrate the information flow and influences and constraints from its surrounding environment, and MATLAB/Simulink is used in virtual development and testing of controller models and electronic/mechanical models. In comparison to mechanical systems, however, model-based development and testing of software-based systems in the automotive industry is in its infancy. For example, because of the lack of a formalized modeling language, [2] point out that modeling is only applied to certain steps in PD, yielding inefficient exploitation of its possibilities. Another issue considers the lack of appropriate digital tools that support a holistic approach to model-based development, where insufficient integration possibilities such as linking engineering data to models and compatibility between different tools are examples of issues that need to be resolved. Finally, [2] highlight the necessity of cost efficient use of modeling by reducing the probable increased workload of maintaining and building the models. This was also brought up by some of the interviewees.

The above discussion indicates that it is not yet feasible to integrate manufacturing processes in model-based development since there are still issues that must be resolved in terms of the modeling of the software-based systems themselves. Nevertheless, despite the challenges of model-based development, it enables earlier and more frequent prototyping loops, which facilitates cross-functional lean engineering practices such as genchi gebutsu. To develop models and tools that can be used by the manufacturing function, it is recommended that the five sub-principles in the selection of tools and technologies in principle 11 be considered, as described in [9]. Moreover, manufacturing engineers should be involved in the further development of modeling tools and working procedures since it is preferred that this development be an evolving process rather than a big bang. In practice, it may start with an elaboration of work procedures comprising analyses based on observations of computerized visualizations of the software-based systems on an appropriate abstraction level (e.g. the functional level) and by using cross-functional techniques such as FMEA, DFMEA and elaborated checklists. The systems for digital development of software-based systems, manufacturing processes and surrounding entities should also be seamlessly integrated, which would facilitate cross-functional access to the necessary data and applications.

Issue 6 Lack of SW competence in the manufacturing organization & Issue 7 Too little knowledge about the manufacturing processes in the PD organization leads to difficulties in introducing new products in production: A product engineer at Toyota typically starts his or her development with basic training such manual work of building cars and customer interaction at dealers during the first year. In the second year they are assigned to an improvement project with the objective of training and giving new technicians an understanding of the Toyota engineering culture. Next, the rookie engineers are assigned to an engineering specialty where they undergo an intensive OJT period. Unlike the situation in other companies where product engineers seldom stay at the same engineering tasks, [9] argue in principle 7 that the lean company engineers stay and develop their technical capabilities and an in-depth and understanding of the PD process at the same specialty for many years (approximately eight years). Since Toyota views itself as an automotive manufacturing company, it emphasizes the importance of achieving manufacturable designs and sees manufacturing engineering as a core competence in the
An Investigation of Manufacturing Challenges in Automotive Software Development from a Lean Perspective

to be able to technically challenge PD engineers, the recruitment and development of manufacturing engineers are equally comprehensive and thorough as for design engineers at Toyota. In contrast, other companies believe that manufacturing engineers do not require the same level of technical capabilities as product engineers since their work is less professional.

In principle 9, [9] claim that companies that attempt to implement a lean system often just copy the explicit knowledge of lean tools and techniques and overlook the need to tap into the tacit knowledge of a lean culture. Nelson and Winter [18] suggest that tacit knowledge is embedded in the skills of workers and in work routines and shared understandings that, in combination, comprise an organization's distinctive capabilities. Lean enterprises understand the importance of developing, diffusing and maintaining tacit knowledge in the organization by recognizing it as a natural element in daily work. In accomplishing this, [9] discuss the importance of creating learning networks facilitating knowledge transfer from PD to manufacturing. Argot and Ingram [19] define knowledge transfer as "the process through which one unit (e.g. group or department) is affected by the experience of another". Further, [19] state that there is a growing agreement that organizational knowledge explains the performance of organizations.

An adapted network model emphasizing a formal structural approach is presented in Figure 3, to analyze the capability of the organizations studied to create learning networks. The model has been derived from organizational charts and interviews at the organizations studied. To facilitate a comprehensive understanding of the patterns of interrelationships in the area of interest, the ties have been clustered to an overall unitary abstraction level that is representative for both companies. Further, to distinguish between operative and managerial work, the network model has been divided into an operational and a managerial area. The model discloses a clustering of manufacturing actors on one side and PD, purchase, suppliers and product planning on the other side in the operational area (here called the PD cluster). Another observation in the operational area is the single tie between the clusters of PD and manufacturing passing through the Manufacturing Engineering (ME) unit. This poses that the ME unite possesses a so called broker role according to one of the six analytical principles considering network analysis provided in [20]. The network model clearly indicates a low dense network between the PD and manufacturing clusters where the ME unit has a high degree of betweenness centrality. Meyer and Rowan [21] suggest that, as density increases, shared norms and behavior tend to diffuse across the network. Hence, it may be possible to anticipate that the issues discussed here can be related to the low density between the manufacturing and the PD cluster and the broker role of ME, impeding the propensity of the organizations to transfer knowledge between each other.

To enhance the knowledge of manufacturing processes among designers and increase the SW competence in manufacturing there seems to be three recommendations that can be derived from to the LPDS model. First, it is recommended that technical capabilities among design and manufacturing engineers should be leveled out by equalizing the conditions for recruitment and development. The second recommendation pays attention to manufacturing capabilities when recruiting and developing design engineers. Finally, it is recommended that the ability to create learning networks in the formal organization be considered. However, to implement appropriate improvements facilitating the diffusion of knowledge between the organizations it must be further investigated how knowledge is transferred through informal processes and communication channels on deeper structural levels.
4. CONCLUSIONS

This paper provides an extensive investigation of the results of an explorative multiple case study focusing on the interaction between manufacturing and PD in automotive software development by using complementary perspectives found in related literature. A review of pertinent literature related to the research area acknowledge that an overall view among researchers, e.g. [4] and [5], is that the automotive industry pursue to simultaneously achieve product differentiation on a broad market and cost leadership by adopting a lean approach to manufacturing and PD. Consequently, the analyses and discussions of issues identified in the multiple case study conducted are based on the lean concept and inherent principles.

The analyses indicated that it was possible to establish linkages between issues primarily focusing on the area of interest and lean principles and practices leading to feasible recommendations that form a framework for further implementation of improvements. An overall conclusion is that it is useful to adopt a lean approach within the realm of the investigated area since it seems to constitute a foundation for appropriate remedial measures. Proposed recommendations however need further refinement and assessment in future research.

5. REFERENCES

Reusability Enhancement of Automotive Components through Disassembly Study

D.A. Wahab*, Lily Amelia, A.R. Ismail, R. Ramli, N. Muhammad and C.H. Che Haron

Department of Mechanical and Materials Engineering
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, Malaysia

ABSTRACT

Automotive component reuse as one of the product recovery strategy is now gaining importance in view of its impact on the environment. Research and development on components design and manufacturing as well as tools and methods to facilitate reuse are under way in several countries. To enable reuse, components have to be assessed and its reliability and life time predicted. This paper presents the development work on the prediction of reliability and durability of reuse components using Artificial Neural Network, focusing on one of the determinant factors for reuse namely disassemblability of a locally produced car door. The existing design of the car door was analysed and design changes were proposed to improve disassemblability. Significant improvements to the design and disassembly time is presented and discussed in this paper.

1. INTRODUCTION

Environmental concern and government legislation have forced manufacturers in many countries to consider product life cycle issues and deal with product recovery at the end of product life cycle. The automotive industry is one of the leading industries in this environmentally conscious manufacturing and product recovery.

Product recovery strategies include reuse, remanufacture and recycling in which reuse is the highest hierarchy in these recovery strategies (see Fig. 1). In reuse, products or components that have not undergone excessive deterioration in retirement phase can be recovered, possibly after being regenerated through immediate processes as components for reassembly with savings in energy, possible emissions, costs related to the process of producing the parts and in volumes of virgin materials [1].

In USA, 95% of cars and trucks that are retired each year go the recycler and for each of those cars, 75% by weight is recovered for reuse [2]. In the European Union countries, the European Union End of Life Vehicles (ELVs) Directive has passed laws to the member countries to reuse and recover 85% by weight of the average vehicle by the year 2006 and this percentage increases to 95% by the year 2015 [3].

In Malaysia, the establishment of Proton in 1985 and subsequently Perodua in 1993 acted as a catalyst to the development of the automotive sector in Malaysia including the development of local automotive components manufacturers. Currently, there are 4 passenger and commercial vehicle manufacturers, 9 motor vehicle assemblers and 343 components/parts manufacturers in Malaysia with total production is 441 678 vehicles in 2007 [4]. The government policy provides support and incentives to promote a competitive and viable automotive sector and to enhance local capabilities in the automotive sector. However, the establishment of local automotive industries that have environmentally consciousness is still one step ahead and need support from government and automotive manufacturers.

Implementing reuse as one of product recovery strategies is conditioned by a wide range of factors that determine its effectiveness. Giudice et al. [1] stated that the determinant factors of reuse include physical and mechanical deterioration, reliability and durability of components and system, ease of disassembly and technological applications.

* Corresponding Author: Tel: +603-89216118, Fax: +603-89259659, Email: dzuraida@eng.ukm.my.
obsolescence. From this point of view, it is important to design products that have high durability and reliability as well as easily disassembled to facilitate reuse.

Design for disassembly as a part of design for reuse is expected to improve the ease of disassembly of product or component. Disassembly is defined as a systematic removal of the desired parts from an assembly with condition that this disassembly process does not lead to any damage to the parts [5]. It differs from dismantling due to its non destructive and semi destructive (irreversible) characters. Amezquita et al. [6] stated that remanufacturable assemblies should be designed with following characteristics: ease of disassembly, ease of reassembly, ease of cleaning, ease of inspection, ease of part replacement, more reusable components and standardization of modular components, fasteners and interfaces. The selection of fastening and joining method as well as disassembly and reassembly method will influence the disassembly time and hence the disassembly cost [7]. Gupta and Mc Lean [8] also stated that there are many advantages of designing disassembly including reduction of work needed, uniformity and predictability in product configuration, simple and rapid disconnecting operations, ease in separation and treatment of removed parts and reduction in product variability.

Currently, this study is focused on the improvement of disassembly design of car door components as an initial stage in improving the reliability and durability of the car door for reuse. Through this study, an existing design of a locally manufactured car door was examined and improvements in disassembly design were proposed and analysed.

An artificial intelligence technique such as artificial neural networks (ANNs) can be used to predict the reliability and durability of automotive components. In recent years, such method has been widely applied in many areas of decision making including product design. A neural network can be defined as a model of reasoning based on human brain [9]. The ability of ANNs to learn and understand the problem and improve its performance through learning process can be applied to predict the life time and reliability of automotive reuse component based on its disassembly and material design.

2. METHODS OF STUDY

Seven criteria for reusability were used to evaluate the disassembly design of the car door including ease of disassembly, ease of reassembly, ease of cleaning, ease of part replacement, components/parts reusability,
manufacturability and manufacturing cost. These criteria and the weightage of criteria were identified based on literature and an interview with a local car manufacturer.

Several design concepts were then generated based on literature and interview with a local car manufacturer. These design concepts were compared with the original design using Pugh concept evaluation matrix [10] and the best design was chosen based on those seven criteria. A score 1 to 5 is given for each criteria in comparison with the datum (the original design). Rating 1-5 is represented as: 1- very bad, 2 – worse than, 3 – same as, 4 - better than and 5- very good. The concept that has the highest weighted sum of score value is chosen as the best design concept.

The disassembly time of the proposed car door was determined to ascertain whether this new design is improved as compared to the original design. A neural network model is also proposed for further analysis in determining the reliability and durability of the car door.

3. RESULTS AND DISCUSSION

3.1. EXISTING CAR DOOR DESIGN: ANALYSIS OF DISASSEMBLY

In this study, the right hand side (RHS) of a car door produced by a local manufacturer is shown in Fig. 2. In the original design, the door skin is crimped to the door frame around its sides except the upper side. There are two spot welds on the front side. These crimping and spot welding cause difficulty in removing the door skin from the door frame. Disassembly process of this original design may cause damage to the door skin and the door frame. In addition, the disassembly process will require a long disassembly time. In the original design, the door frame comprised of two parts that are connected using spot welding. This weld is the weakest part of the door frame and has high damage risk. The other part such as door sash is connected to the door frame using spot welding. The outer window panel is connected to the door frame using both snap fit and screw, while the inner window panel uses clip.

![Figure 2: Right hand side door of a locally manufactured car](image)

3.2. REDESIGNING OF THE CAR DOOR: DESIGN CONCEPTS GENERATION AND SELECTION

In this study, a morphology chart was used in the generation of design concepts as presented in Table 1. In the morphology chart, the idea is gathered based on the sub functions of the car door. There are four important sub functions of a car door such as to hold depression, to hold impact from the outside, to set window glass position and to prevent water inlet. Different types of fasteners are proposed for each concept to satisfy these sub functions and the criteria for reuse. For example, to hold the door skin to the door frame, several options of fasteners are proposed such as crimping (concept 1 and 2), bolts and nuts (concept 3), rivets (concept 4) and screw (concept 5). Meanwhile, the door skin is fixed to the door frame using options that include no groove (concept 1 and 2), one groove (concept 4) or two grooves (concept 3 and 5). Concept 1 and 2 propose the design of the main door panel as two entities whilst concepts 3-5 uses a one entity door panel. Meanwhile, door sash can be designed with a sharp edge groove or a circular edge groove in order to drive the window glass movement. In concept 1 and 5, screw is used to fix the outer window position, while clip is chosen for concept 2 and 4, and snap fit for concept 3. To fix the window glass position the door sash can be either crimped (concept 1 to 2) or spot welded (concept 3 to 5) to the
main door panel. For the sub function of preventing water inlet from the outside, the inner window panel position can be fixed using screw (concept 1 and 5), snap fit (concept 2 and 3) and clip (concept 4).

Using the Pugh Concept Selection Matrix as shown in Table 2, concept 3 obtained the highest score for ease of disassembly, ease of parts replacement and reusability of parts/components criteria. The choice of fasteners in this concept such as bolts, nuts and snap fit improves the disassembly process compared to the use of other types of fasteners such as crimping and spot welding. Consequently, this will increase the reusability of parts or components. However, the use of bolts and nuts will pose difficulties on reassembly compared to crimping. Concept design 3 has also a lower manufacturability and a higher manufacturing cost compared to the original design. However, as a whole, concept design 3 was found to be the best redesign concept with the highest score of 3.65 and is proposed as the new design for the car door.

### Table 1: Design concepts of car door for reuse

<table>
<thead>
<tr>
<th>Sub Function</th>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
<th>Concept 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 To hold depression</td>
<td>Crimping</td>
<td>Crimping</td>
<td>Bolt and nut 2 grooves</td>
<td>Rivet 1 groove</td>
<td>Screw 2 grooves</td>
</tr>
<tr>
<td>1.1 To hold door skin</td>
<td>No groove</td>
<td>No groove</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 To fix door skin position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0 To hold impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Main door panel</td>
<td>2 entities</td>
<td>2 entities</td>
<td>1 entity</td>
<td>1 entity</td>
<td>1 entity</td>
</tr>
<tr>
<td>3.0 To set window glass position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 To drive window glass</td>
<td>Sharp edge groove of door sash</td>
<td>Sharp edge groove of door sash</td>
<td>Circular edge groove of door sash</td>
<td>Circular edge groove of door sash</td>
<td>Circular edge groove of door sash</td>
</tr>
<tr>
<td>movement</td>
<td>Screw</td>
<td>Clip</td>
<td>Snap fit</td>
<td>Clip</td>
<td>Screw</td>
</tr>
<tr>
<td>3.2 To fix outer window</td>
<td>Door sash and main door panel are cramped</td>
<td>Door sash and main door panel are cramped</td>
<td>Door sash and main door panel are spot welded</td>
<td>Door sash and main door panel are spot welded</td>
<td></td>
</tr>
<tr>
<td>panel position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3 To fix window glass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0 To prevent water inlet</td>
<td>Screw</td>
<td>Snap fit</td>
<td>Snap fit</td>
<td>Clip</td>
<td>Screw</td>
</tr>
<tr>
<td>4.1 To fix inner window</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>panel position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Pugh concept selection matrix

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Weight</th>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
<th>Concept 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>score</td>
<td>total</td>
<td>score</td>
<td>total</td>
<td>score</td>
<td>total</td>
</tr>
<tr>
<td>Ease of disassembly</td>
<td>0.20</td>
<td>1</td>
<td>0.20</td>
<td>1</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>Ease of reassembly</td>
<td>0.20</td>
<td>4</td>
<td>0.80</td>
<td>5</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>Ease of cleaning</td>
<td>0.10</td>
<td>1</td>
<td>0.10</td>
<td>1</td>
<td>0.10</td>
<td>3</td>
</tr>
<tr>
<td>Ease of part replacement</td>
<td>0.15</td>
<td>1</td>
<td>0.15</td>
<td>1</td>
<td>0.15</td>
<td>5</td>
</tr>
<tr>
<td>Reusability of components</td>
<td>0.20</td>
<td>1</td>
<td>0.20</td>
<td>2</td>
<td>0.40</td>
<td>5</td>
</tr>
<tr>
<td>Manufacturability</td>
<td>0.10</td>
<td>4</td>
<td>0.40</td>
<td>4</td>
<td>0.40</td>
<td>1</td>
</tr>
<tr>
<td>Manufacturing cost</td>
<td>0.05</td>
<td>5</td>
<td>0.25</td>
<td>4</td>
<td>0.20</td>
<td>2</td>
</tr>
<tr>
<td>Total score</td>
<td>2.10</td>
<td>2.45</td>
<td></td>
<td>3.65</td>
<td></td>
<td>3.40</td>
</tr>
<tr>
<td>Ranking</td>
<td>5</td>
<td>4</td>
<td></td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
3.3 THE PROPOSED NEW DESIGN OF THE CAR DOOR

In this new design, instead of using crimping the attachment of door skin to the door frame uses the “tongue and groove” principle. This concept was proposed by Amezquita et al. [6] and is modified by sliding the door skin from the front side. There are 2 grooves and 2 tongues to guide the door skin direction into the door frame (see Fig. 3). Bolts and nuts are used to hold the door skin. Four bolts (two at each side) are used to fix the position of door skin to the door frame. Compared to the original design, this new design allows an easier disassembly of door skin and door frame and prevent damages during the disassembly process. This also leads to reduction in disassembly time and recovery cost. The new design will provide more opportunity for reuse for the car door parts such as the door frame and door skin.

Figure 3 : Tongue and groove principle in new design

It is also suggested that the door frame be constructed as one entity instead of two separate entities connected by spot welds (see Fig. 4). This will increase the possibility for reuse because spot welding connection is not durable. This design will enhance the capability of the door frame to hold impact from the outside.

Figure 4 : One entity of door frame

As similar to the original design, the connection between door sash and main door panel is maintained using spot welding. The door sash is proposed to have a circular edge groove as the circular edge groove does not trap dirt and is easier to be cleaned than a sharp edge groove (Fig. 5).
The outer window panel is designed using snap fit to replace the use of both snap fit and screw (see Fig. 6). This new design enables ease of disassembly and reassembly compared to the original design. The snap fit was designed for easy removal only from the inside, therefore vehicle security is still maintained. It is also proposed that the inner window panel uses snap fit instead of clip (see Fig. 7). Even though clips are able to hold strongly, they are not easy to disassemble. Snap fit is easier to disassemble and reassemble compared to the clip.

3.4. THE ANALYSIS OF DISASSEMBLY TIME

The proposed redesign enables the car door to be disassembled easily and therefore will reduce disassembly time as well as recovery cost hence enhancement to the reusability of the car door. Based on Boothroyd [11], the disassembly time of the proposed design of car door is estimated around 330 s or 5.5 min (see Table 3). It is much lower than the disassembly time of the original design.
### Table 3: Estimation of disassembly time of the new design

<table>
<thead>
<tr>
<th>Component/Material</th>
<th>Quantity</th>
<th>Preparation time (s)</th>
<th>Disassembly time (s)</th>
<th>Total disassembly time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer window panel</td>
<td>1</td>
<td>5</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Bolt</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Bolt and nut</td>
<td>1</td>
<td>5</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Adhesive glue</td>
<td>-</td>
<td>180</td>
<td>-</td>
<td>180</td>
</tr>
<tr>
<td>Door skin</td>
<td>1</td>
<td>20</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td><strong>Total disassembly time</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>330</strong></td>
</tr>
</tbody>
</table>

The disassembly time of the proposed design includes the time for disassembly of the outer window panel, removing all bolt and nut fasteners that connect the door skin and the door frame, removing the adhesive glue that is put on the door skin surface and removing the door skin from the door frame. The original design of the car door is estimated to have the same disassembly time as the Chrysler left hand side door. This Chrysler door has similar disassembly design with the car door original design that also uses crimping and spot welding between the door skin and the door frame. The estimated disassembly time of Chrysler car door is around 39.5 min [6].

### 3.5. RELIABILITY AND DURABILITY PREDICTION USING NEURAL NETWORK

Artificial neural networks (ANNs) has been widely used in recent years due to its ability in providing satisfactory and acceptable solution for complex problems in which the conventional mathematical methods are not able to solve. ANNs are inspired by the human brain that consists of a number of interconnected processors, called neurons. The neurons are connected by links and each links has a numerical weight associated with it. A neural network ‘learns’ through repeated adjustments of these weights. A multilayer neural networks is a feed forward neural network that consists of one or more hidden layers. Typically, the network consists of an input layer of source neurons, at least one middle or hidden layer of computational neurons and an output layer of computational neurons (Fig. 8).

![Back propagation neural network](image)

Figure 8: Back propagation neural network

A lot of training algorithm are available, however the most popular method is back propagation [9]. In a back propagation neural network, the learning algorithm has two phases. First, a training input pattern is presented to the network input layer. The network then propagates the input pattern from layer to layer until the output pattern is generated by output layer. If this pattern is different from the desired output, an error is calculated and then propagated backwards through the network from the output layer to the input layer. The weights are modified as error is propagated. The output of neurons in the hidden layer are calculated by:

\[
Y_j(p) = f \left[ \sum_{i=1}^{n} x_i(p) \times w_{ij}(p) - \theta_j \right]
\]

\[\text{................................. (1)}\]
where $f$ is the transfer function, $n$ is the number of input neuron $j$ in hidden layer, $x_i(p)$ is the input neuron value at iteration $p$ and $w_{ij}(p)$ is the weight and $\theta_j$ is a threshold applied to the neuron. As well, the actual output of neurons in output layer are calculated by:

$$Y_k(p) = f \left[ \sum_{j=1}^{m} x_{jk}(p) \times w_{jk}(p) - \theta_k \right]$$  \hspace{1cm} \text{(2)}$$

where $m$ is the number of input of neuron $k$ in the output layer. Tansig and purelin transfer functions will be used in the proposed model. Both are very common choices for multilayer neural network. Some back propagation learning algorithms such as Levenberg-Marquardt, Resilient Back-propagation and Scaled Conjugate Gradient will be compared to obtain the best results.

The data requirements for the proposed neural network model will be derived from finite element (FE) analysis using computer aided engineering software. By simulating loads, disassembly form and type of materials, the proposed design of car door for reuse will be analysed using FE method so that various stress and cycle to failure could be determined. Based on this, ANNs will predict the durability and the reliability of the car door. Work on durability and reliability of the proposed design of the car door is currently being carried out. The design improvement of the car door for reuse is not only to improve disassembly but also to increase durability and reliability of the car door components.

4. CONCLUSIONS

Disassembly design is one of the determinant factors in the enhancement of reusability of products at the end of its life cycle. This study has been successful in proposing a new design of car door for reuse by pursuing an ease of disassembly design of the car door. Some design modifications to a locally manufactured car door has been carried out to enable ease of disassembly and reassembly leading to a significant reduction in disassembly time. This initial study is necessary for further research in enhancing reuse of the automotive component.

ACKNOWLEDGEMENTS

The authors would like to thank to the Ministry of Higher Education of Malaysia for sponsoring this research under Research University Grant UKM-GUP-BTT-07-25-149.

REFERENCES

A CAD System for Extraction of Mating Features in an Assembly

Arun Tom Mathew¹, C.S.P. Rao²

¹Lecturer, School of Mechanical Sciences, Karunya University, Karunya Nagar, Coimbatore, Tamilnadu – 641 114, India
²Professor, Mechanical Engineering Department, National Institute of Technology, Warangal, Andhra Pradesh – 506 021, India

ABSTRACT

The information and knowledge about a product and its assembly are necessary to generate all feasible assembly sequences of that product. A mechanical assembly is a set of interconnected parts representing a stable unit in which each part is a solid object. Assemblies contain a very large amount of information and complex relationships. Surface contact between parts reduces the degree of freedom for relative motion. Identifying assembled parts as well as their contact surfaces is very important in design and manufacturing since this information is necessary. The problem is to not only make the information available but also use the relevant information for making decisions, especially determination of the optimum assembly sequence. This paper describes a system for processing assembly models and extracting assembly related data. This data is used to identify the relationships between different components of an assembly thus encouraging generation of feasible assembly sequences. The system works on the assumption that the designer explicitly defines joints and mating conditions and such relations can be extracted to form a signature of the assembly.

1. INTRODUCTION

Assembly is one of the most important activities in the manufacture of a product because of its complex nature. More than 30% of total industrial product labour costs [1] and 50% of product manufacturing cost are attributed to the cost of the assembly [2]. The assembly process consists of a number of different stages, such as putting together all the parts and subassemblies of a specified product. Most mechanical products can be assembled in several ways, meaning that different sequences of assembly operations can result in the same final product. Each such sequence implies a different degree of difficulty for the various assembly operations, resulting from different mechanical constraints imposed by the different sequences of operations. Selection of a good sequence of assembly operations is a crucial factor in maximizing the production profitability and has great impact on the assembly line balancing, machine utilization and feasibility of subassembly operations. Numerous assembly sequence generation methods have been developed to systematically explore all the feasible sequences. In addition to the geometric and topological information, the precedence constraints and inter-part relationships are required to study the assembly. But the above information needed is not openly available. In assembly design, when two parts are assembled together, some surfaces of the two parts will come in contact with each other. The contact area results in a relationship in an assembly [3].

2. RELATED WORK

Many researchers have attempted to generate and evaluate the assembly sequences for a product. Bourjault proposed a procedure which obtains all the precedence knowledge about the liaisons of an assembly by answering a set of structured questions based on his proposed liaison model of the assembly. De Fazio and Whitney simplified Bourjault’s procedure and reduced the number of questions to be asked to 2n against 2^n of Bourjault’s [4]. These two methods study the assembly from the point of view of assembling the product. For the representation of the precedence knowledge for an assembly, there have been several methodologies widely used in the past such as set

¹ Corresponding author: E-mail: arun.tom123@gmail.com
theory, binary matrix, directed graph, establishment conditions and precedence relationships. However, these methods can only represent partial assembly precedence knowledge. The problem with the above methodologies is that they are difficult to use to generate detailed assembly plans automatically and to deal with the coordination and feasibility of various subassemblies efficiently. An assembly can have many different feasible assembly sequences. As it is difficult to represent each sequence individually, it is necessary to design a method to represent all the sequences in an efficient and compact manner. Gottipolu and Ghosh developed a method for automatically generating assembly sequences from a solid modeller that generates two types of matrices algorithmically [5]. The aim of extracting inter part relations from an assembly design is to support automatic assembly sequence generation. According to literature, the inter part relation have been defined in terms of the spatial relationships [6], degree of freedom [7], liaison relations [8], hierarchy relationship, mating types [9]. Instead of total human interpretation of the assembly design, a direct CAD database interface approach has been proposed to extract the relation with minimal manual involvement.

3. MODELLING OF ASSEMBLY

Assembly Modelling is an extension of geometric modelling that facilitates that construction, modifications and analysis of complex assemblies. A product can be considered to be an assembly of elementary components. In assembly modelling, a product model is created representing a product consisting of several smaller components. Parts and components are added to an assembly by specifying mating conditions or constraints. Because of these smaller components, the focus in assembly modelling will be not only on these components, but also on the relations between these components. To describe a product, the elementary components and the relationships between them must therefore be defined. A component that cannot be subdivided into smaller components is called a single part. A group of components merged together is called an assembly. Decisions made during the creation of a model can have great impact on the complete life cycle of the product. Assembly problems are generally handled in graph form. The description of the relationship can be generated manually or automatically if we have the CAD model of the components. The assembly design process along with the top level breakdown of assembly form the backbone for the generation of assembly information.

4. GENERATION OF ASSEMBLY RELATIONSHIPS

SolidWorks, the commercial CAD system, is used as the main feature-based design environment. The benefit of using SolidWorks is that it includes a complete API (Application Programming Interface) with functions that can be called from Visual Basic. In addition, SolidWorks shares the same solid modelling engine as Unigraphics and several other CAD systems like the Pro/Engineer and Catia. Together, these CAD systems account for large user and application bases.

The description of the relationships among the features of various parts is required for an assembly component. These features can be classified into assembly features and primal features. It is the primal features that participate in assembly constraints. The assembly module automatically determines which relationship is meant by the user based upon the features involved in the relationship and updates the degrees of freedom accordingly. The primary mating conditions are align, mate, mate entity, align offset, insert, orient etc. The align condition requires that the axial centre lines of two parts be collinear. The mate condition requires that the two mating faces lie in the same plane with their outward normal opposing each other. The offset condition requires that the two faces lie in parallel planes with their outward normals in the same direction.

The relationships between a pair of parts are specified by the user in terms of their features and the mating conditions between them. The individual parts in an assembly are created before the assembly module is invoked. The assembly modelling module requires information about the relationships between the part features. The
information specified for each mating condition includes the ID of the mating feature and the type of mating conditions.

To build a list of all the characteristics of an assembly the assembly format is developed to store all the characteristics in an assembly as its signature. The method explores the assembly tree in depth. While exploring the assembly hierarchy, it extracts assembly related information for each part.

a) The method retrieves the constraints and dimensions used to specify the position of the part.

b) It identifies which entities are used to constrain the part or subassembly.

c) It identifies the parent features and part of each geometrical entity in use.

The assembly information generated is represented in an object oriented way to generate assembly strategies.

5. MATE INFORMATION

The Application Programming Interface (API) functions used in this paper are SolidWorks functions. The API functions are essential for developing the application software. The names of the mate features, the types, identities and the types of the mate surfaces, the mate clearances and the reference features etc are included in the mate information. In SolidWorks there are the following objects related to the mate information like Mate, Mate Feature, Face and Surface. The Mate Object allows access to various assembly mate parameters. The MateEntity object enables access to mated objects and the assembly mate definition. The Feature Object allows access to the feature type, name, parameter data and to the next feature in the FeatureManager design tree. The Face Object allows access to the information of surfaces related to the mate. The Surface object provides a function that enables to get the surface type and various surface definition data, as well as, evaluate and reverse evaluation locations on the surface. We gain access to the mate information through these objects in the FeatureManager design tree.

In SolidWorks, the current features of assembly bodies are obtained by traversing the FeatureManager tree. An API function of the Mate object “GetMateEntities” is used to get the mate entities related to the current mate. SolidWorks defines the mate relationships like perpendicular, tangent, coaxial, parallel, distant, angular, symmetric etc. which can be gained by means of an API function of the feature object – “GetTypeName”. The face identities related to the mate entities can be got through an API function of the Face Object called “GetFaceID”.
To illustrate the process, an example consisting of 9 parts of a nut cracker is given. The nut cracker taken for the present study has 9 components namely the Base, Cylinder, Hinge, Piston, Adjuster base, Adjuster, Connection, Handle and Pin. SolidWorks software was used to model the assembly. The individual components were created as separate geometric models in the part models in the part mode and saved as “.sldprt” files. Next, the assembly modelling mode is invoked and the Base is taken as the support component. After specifying the assembly constraints, the assembly was built by adding the remaining components to the base part. Adding to the Base component are three sets of components namely the Adjuster base, Cylinder and the Hinge. The Adjuster base acts as a sub assembly consisting of the Adjuster. The Cylinder acts as a subassembly having the piston, two connectors and pin as its subcomponents. Further the hinge subassembly consists of the pine and handle which in-turn are connected to the cylinder subassembly via the connectors. All the components are assembled using the mate attributes like the Coincident, Parallel, Perpendicular, Tangent, Concentric, Distance and Angle. The completed assembly model is then saved as a “.sldasm” file.
The mate option is used to assemble the faces of two components and the align option is used to align the axis of the two components. Select the faces, edges, planes and so on that is to be mated together. All the mate types are always shown in the PropertyManager, but only the mates that are applicable to the current selections are available. “Coincident” option positions selected faces, edges, and planes (in combination with each other or combined with a single vertex) so they share the same infinite plane and positions two vertices so they touch. “Parallel” option places the selected items so they remain a constant distance apart from each other. “Perpendicular” option places the selected items at a 90° angle to each other. “Tangent” option places the selected items tangent to each other (at least one selection must be a cylindrical, conical, or spherical face). “Concentric” option places the selections so that they share the same center line. “Distance” option places the selected items with the specified distance between them. “Angle” option places the selected items at the specified angle to each other. The final assembly resulting from satisfaction of all the mating conditions is shown in the Figure 3.
Figure 4 is the assembly’s structural view and considering the relevant geometric relations and assembly joints in the assembly, a relation graph can be constructed. Figure 5 shows the state of the assembly showing the base component and the sub assembly. A mate graph for the assembly can be constructed as shown Figure 6.

![Diagram showing assembly components](image)

**Figure 5: Base Part (BP) and Subassemblies**

![Diagram showing mate connections](image)

**Figure 6: Mate Diagram of various connections of the assembly**

Figure 7 shows the exploded view of the assembly having the 9 components. The body base is taken as the base component and the subcomponent consists of the Cylinder, Hinge, Piston, Adjuster base, Adjuster, Connection, Handle and Pin.
7. CONCLUSION

In this paper, a system for processing assembly models and extracting assembly related data is described. The utilization of mating features for assembly modelling is important, because through these features it is possible to identify the information necessary to perform assembly analysis. The information is used to generate a list describing the links between the assembled parts, the involved features and the type of link explicitly to facilitate assembly analysis and planning. These representations allow the exchange of design intent and assembly constraint information between modelling, analysis and planning systems. The aim of developing such a scheme is to reduce human interaction in the process of creating assembly plans.

REFERENCES

System Based on Frame to Formalize and (Re)use Functional Knowledge in Functional Modeling of Parts

1Francisco das Chagas Mendes dos Santos, 2Altamir Dias

1Federal Institute of Education, Science and Technology of the Amazonas - IFETAM
fcomsantos@emc.ufsc.br

2CAD/CAM Laboratory, Department of Mechanical Engineering, Federal University of Santa Catarina - UFSC
altamiar@emc.ufsc.br

ABSTRACT

On the methodologies of designing a product/part, on the “analysis activity” in the stage of conceptual design, design/designers staff develops many ideas and functional knowledge about the solution of the trouble in design through an explicit/implicit way in the function textual description on natural language (FTDNL). However, such ideas and functional knowledge are arduous formalized and stored for effective (re)use in designs of the same, or others, application fields. Therefore, the main intent of this assignment was to design a system based on frame to formalize and store explicit/implicit functional knowledge into the part’s FTDNL – named SISFCO (AL). Next, took place syntactic (or grammatical), semantic, and pragmatic analysis of every FTDNL. The SISFCO (AL) is composed of three main modules: 1) FTDNL editor; 2) Natural Language Processing System (NLPS) and 3) Frame System. The FTDNL editor is based on a function model that upholds linguistic approach. The function model is based on the linguistic approach and is supported by a SPLN. The SPLN is composed by syntactic, semantic and pragmatic analysis modules and the speech one. Analysis Module a) Syntactical: analyses the grammatical sort of each FTDNL constituent. b) Semantic is assisted by pragmatic analysis on the settlement of every constituent’s meaning. c) Speech assists the former modules on the arrangement of the constituents of the meanings associated to the same contexts. Finally the frame system organizes the meanings of constituents, functional information, related to the functional knowledge of the part’s design context. The main contributions of the methodology based on linguistic approach are: i) possibility of formalization and storage of functional knowledge, ensemble of functional information, explicit/implicit in the FTDNL and ii) effective (re)use of functional knowledge during the part’s functional modeling.

1. INTRODUCTION

In the literature about methodology of design, consensually the functional knowledge plays an important role in the designing of a product or part. This importance comes from the need of setting up conceptual solutions through the functional structures of the product (system), subsystems and parts.

To Roozenburg and Eekels [1], a product can be planned to be a model of behavior composed of an ensemble of functional structures of subsystems and parts. Such functional arrangements that form the product must be observed as constituents (non-materially defined) apt to do their global function. However, to idealize a model of behavior to the product or part, the designer/design-staff needs to organize, compatibly, an ensemble of behaviors carried out by geometries and subjects of geometrical details of an ensemble composed of physical structures of parts. To guide the idealization of the behavior model to the product or part, the designer/designers-staff working out the design describes textually the requirements/restrictions for space and design (design specification of the product – PSP) of the properties r required by the customers a nd others kn own by them so t hat the product r ealizes its functional purpose (un)desired.

Roozenburg and Eekels emphasized that the idealization of a model for the product’s behavior are not only applications of scientific knowledge, but, especially, an idealized and structured representation of known behaviors of geometrical details (GDs) of an ensemble composed of physical structures of parts, in which these elements characteristics and their inter-relations are qua altitatively defined. Due to that, the core of the problems on the methodologies of design lies on the cataloguing of the functional knowledge about the causal model (related to the
physical, chemical, biological effects), geometries and the manner and operational conditions of its uses explicit/implicit on the textual description of function on natural language (TDFNL). Thus, the cataloguing of functional knowledge included on TDFNL comprises the register of semantic values of entities and attributes of design’s aspects and process expressed by constituents of the TDFNL itself, such as verb, adverb, abstract, concrete, simple or compost noun, adjective, among others and their combinations. Besides, in the formalization of functional knowledge, the registrations of semantic values for the TDFNL’s constituents are guided by the functional points of view of the designer/design-staff and by the domination of the design’s application context.

2. RELATED WORKS

According to Back [2], on the process of designing a product, a principle is basic; it’s the information reforming process, where the basic elements in the arrangement of conceptual solutions can be organized as shown in Figure 1. In this Figure, the basic unit functional information (buFiS) are added by the designer/design-staff working out the design and need to be assisted by some support device to the effective (re)use of buFiSs and must be transformed by some representation way of TDFNL functional knowledge that maintains the designer/design-staff point of view and the domain of the application context.

![Figure 1: General Structure of the buFiS transformation process at the on the process of designing a product with its essential elements](image-url)

In Figure 1 we can observe that the key to the process the designer/design-staff buFiS transformation process is the way how knowledge is included into TDFNL. According to researchers like [3], [4], [5], [6], [7], and [8] there are two types of knowledge which are usually implicit on TDFNL, the: knowledge described by functional purpose and the behavioral knowledge. The knowledge about the functional purpose defines teleological knowledge. The teleological knowledge is the one that enlightens the aim, purpose or finality of the action about the TDFNL. This knowledge can clarify “how” the purpose works when it’s performed, and “when its performing is (un)desired. Some researchers like [9] and [7] even affirmed that defining the functional purpose is describing the intent of the designer’s design intent properly.

The buFiS’ behavioral knowledge about the behavior of the parts GDs and its interactions with another GDs of parts, it’s defined as mapping of the operational conditions thought by designer/design-staff. The operational conditions are an ensemble of behaviors (such as causal process, geometries, feature, fabrication process, among others, when necessary) and attributes (generals or specifics of the parts needed to configure, or formalize, the GDs entities of these parts. According to Fonseca, the general attributes can be basic (for instance, operation, ergonomics, aesthetics, security, reliability and so on) or of the product’s life cycle (such as manufacturability, manageability, usability, maintainability, recyclability, discard etc). Santos e Dias define the behavioral knowledge as an ensemble of buFiS originated from the perceptions of interactions amid functional superficies or interfaces of different GDs in the same part, or amid different ones, under the requirement/restrictions of design and spatial connected to the transformation of the incoming/outgoing entities of energy, material and signal (information).

Santos and Dias [11] proposed a model of representation for the functional knowledge and a methodology to formalize knowledge based on the linguistic approach. Unlike the representation models for classic functional knowledge, such as the Pahl and Beitz [12], which upholds only functional approach, the Santos and Dias one [11] upholds functional and behavioral approach. However, it missed a system guided by a representation of functional knowledge model and a system which uses methodology to formalize and (re)use the buFiS included in the TDFNL.
3. SYSTEM TO FORMALIZE AND (RE)USE FUNCTIONAL KNOWLEDGE ON FUNCTIONAL MOLDING OF PARTS

The intent of this work is to set a system based on a model of representation of functional knowledge of Santos and Dias [11] based on frame to formalize and integrate for effective (re)use of the buFIs of the design and know by the designer/design-staff working out the design according to their point of functional point of view and domination of the application context. On Figure 2 we can see the main interface of the system that supports the formalization of functional knowledge based on the linguistic approach denominated SISFCO (AL). SISFCO (AL) is composed of four main modules: i) module of the editor of TDFNL, ii) module of the natural language processing system (NLPS), iii) module of the frame system and iv) module of the object oriented database management system (OODMS).

3.1. EDITOR MODULE OF PART’S TDFNL OF SISFCO (AL)

The editor of TDFNL of parts is oriented for a function model based on the linguistic approach of Santos and Dias, shown on Figure 3. The function of the TDFNL’s editor is to organize in a hierarchical structure the TD(GF)NL (part’s global function) and its respective TD(PF)NLs (partial functions) and TD(EF)NLs (elementary functions). To formalize an ensemble of buFIs and functional knowledge explicit/implicit on every functionality of the TDFNL, the editor is assisted by a SPLN, frame system and by OODMS. This way, during functional modeling the TDFNL editor of the part can re(use) and (re)modify buFIs stored in other design of parts in real time. Besides, the part’s TDFNL editor can, also, integrate new buFIs of part to be (r e)used or (re)modified in accomplishment time on the same part design.

On Figure 3 we can observe that the part’s TDFNL upholds now the functional approach yet the behavioral one. This kind of support to approaches assists the SISFCO (AL) on the creation of the following reports: i) structures of
TDFNLs with their respective functional purposes. i) structures of the TDFNLs with their respective functional purposes, verbs and explanation of the designer/design-staff in each action of each verb in the part’s TDFNL. iii) structure of the TDFNLs with their functional purposes, causal process and geometries thought by the designer/design-staff to the physical structure of the part. iv) structure of the TDFNLs with their respective functional purposes, causal process and features thought by the designer/design-staff to the design of the part’s physical structure and v) structure of the TDFNLs with every buFIs made available by the designer/design-staff of frame design.

Figure 3: Structure included on an UML of the SISFCO (AL) TDFNL editor module

3.2. NATURAL LANGUAGE PROCESSING SYSTEM MODULE (NLPS)

The NLPS is composed of speech, pragmatic, semantic and syntactic analysis modules. The analysis module: a) Syntactic Analysis Module analyzes the grammatical type of each constituent of the TDFNL, b) Semantic Analysis Module is assisted by the pragmatic one on the determination of the meaning of each constituent and c) Speech Analysis Module assists the former modules at the organization of the constituents of the meanings related to the same contexts.

On SISFCO (AL) the semantic and syntactic analysis of the constituents of each TDFNL are accomplished by CoGOO (http://cogroo.sourfrog.net/index.html). Next, the SISFCO (AL) even uses a post-processor to re-evaluate the constituents of the TDFNL and verify if there are functional syntactic structures. On SISFCO (AL) a syntactic functional structure is an ensemble of constituents contained in a TDFNL which the semantic qualifiers can be related to the design aspects and process defined by the semantic general characteristics proposed by Santos and Dias.

To catalogue the functional knowledge included on the TDFNLs by the designer/design-staff working out the design SI SFCO (AL) uses a class of frame based on how such professionals textually describe the functionalities on natural language. Thus, there’s a correlation between frame and the kind of grammatical structure of the TDFNL. This strategy helps the designer/design-staff to formalize their knowledge included on TDFNL and, at the same time, preserves their functional points of view and their domination of the application context. At the time the designer/design-staff fills the frames with their functional knowledge the speech and pragmatic analysis are simultaneously done.

3.3. FRAME SYSTEM MODULE

The frame module, or class of frames, is derived from several grammatical structures used by the designer/design-staff working out the design on TDFNL. Despite the frame classes uphold any kind of grammatical structure from TDFNL, currently SISFCO (AL) upholds only TDFNL with, at least, a verb and a noun.
Condensing, on SISFCO (AL) the frame systems assist the designer/design-staff on the organization of the meanings of the constituents, buFIs, related to the design/process aspects defined by the general semantic characteristics proposed by Santos e Dias [11].

3.4. OBJECT ORIENTED DATABASE MANAGEMENT SYSTEM MODULE (OODMS)

The OODMS of SISFCO (AL) is db40 (http://www.db4o.com/). The db40 is a databank guided to object of open source and native from Java and Net programming language. As shown on Figure 4, the db40 was chosen due to its objects manager used to consult any kind of object, for example:

a) Analysis of the attributes of the constituents of all the TDFNL on the OODMS;
b) Analysis of the constituents of all the TDFNL on the OODMS;
c) Analysis of TDFNL in all the models of parts on the OODMS;
d) Analysis of (the family of) the parts on the hierarchical structure of the models of parts on the OODMS;

Figure 4: Main interface of the OODMS of SISFCO (AL)

Through Figure 4 we can have a general comprehension about the organization and the sorts of objects on OODMS of SISFCO (AL) starting from the way how the tabs of the db40’s object manager are structured. Examples of the organization and sorts of objects of OODMS of SISFCO (AL) at Figure 4 are:

− Tab Home: are visualized every class instanced on SISFCO (AL), such as, the classes: Frame, Function, Model of Part, Functional Purpose and Slots;
− Tab Frame: It’s had a class Frame there whose objects that are managed by it are the own TDFNL’s frames. The class Frame is responsible for the management of all the buFIs contained on the slots, it is, the buFIs’ facets and their semantic values;
− Tab Function: It’s had a class Function there whose managed object are the global functions, partial or elementary. Every function has the following buFIs: functionality and functional purpose. On the class Function the objects model/part, father and frame are some of the offered by this class for management of the buFIs. For instance, the service related to model/part is made to manage the information about functionality and functional purpose to any part instanced on the model/part. The service father is made to manage the information about the
kind of function, whether global function or partial function. The service frame is made to manage the information about a valid frame, it means, a frame filled with semantic values, according to the designer’s point of view;

- Tab ModeloPeça: It’s had a class ModeloPeça there, whose managed objects are: a) family; b) name; c) image; d) image of the part’s features tree;

- Tab PropositoFuncional: It’s had a class ModeloPeça there, whose managed objects are: description, name and control numeration of the kind of PropositoFuncional;

- Tab Slot: It’s had a class Slot there whose managed objects are word, grammatical class (tag) and attribute. The word is a string of characters (or token) extracted from TDFNL, the grammatical class of the character string is defined by the Normative Grammar of the Brazilian Portuguese Language and the attribute is the semantic value defined by the designer (or system’s user);

4. RESULTS

The result of the research work is the way how the designer/design-staff working out the design organizes the functional knowledge about the problem in the design with the functional knowledge of former experimental design from the same, or another, fields of application contexts. Example of the included formalization on a TDFLN named “Facilitar montagem no pistão” can be seen at the frame shown on Figure 5

5. FINAL CONSIDERATIONS

The main contribution of the methodology based on linguistic approach are: 1) Possibility of formalization and storage of functional knowledge, ensemble of functional information, explicit/implicit in the FTDNL and 2) effective (Re)use of functional knowledge during the part’s functional modeling. The SISFCO (AL) is an important contribution to research line with SPLN applied to the methodologies of product/part. The innovation in the proposed solution lies on the paradigm of linguistic approach, object oriented database management system (OODMS) and the frames system to effectively formalize, store and (re)use functional knowledge, junction of functional information during the moldings of functions in the part.

REFERENCES


Recognition of Interacting Volumetric Features Using 2D Hints

Dušan N. Šormaz*, Chandu Tennety

Department of Industrial and Manufacturing Systems Engineering
Ohio University
Athens, Ohio, 45701-2979, USA

ABSTRACT

Recognition of machining features is an essential step in the development of efficient automated process plans from solid modeling data. This process represents the effective interpretation of the geometric data in a CAD model to create semantically-rich manufacture-oriented features such as holes, slots, pockets and others that may be exploited in downstream CAM/CAPP applications. Most successful approaches towards feature recognition have been based on hint-based procedures operating on a 3D B-Rep model. This paper proposes an approach by which features are identified in a solid model that is built mainly using sweep solid modeling operations. Motivation for this approach is in the fact that sweep operations from 2D sketches are very commonly used approach in the mechanical design process, so the approach may be applicable in practical applications of CAD. Part geometric model is queried for both 2D and 3D geometric elements. Feature hints are generated by an analysis of sweep operations and their 2D sketches, which are defined prior to building the solid model. These hints are then analyzed and validated by applying a two-phase approach: 2D validation in the sketch geometry and 3D validation in the final CSG tree of the solid model. Valid hints are the basis for the creation of a machining feature model that can be input to a process planning module. In addition interaction information for machining features is extracted from both 2D hints and their 3D validation. Feature interaction information is obtained by analysis of face/edge neighborhood and their geometric relations in both 2D and 3D spaces. This approach provides a benefit of performing the majority of geometric analysis in 2D space which is much simpler and computationally more efficient than corresponding analyses in 3D space. Only minimal portion of the analysis is computed ion 3D solid models. The approach is implemented in the Java-based prototype system and is demonstrated and tested on several examples.

1. INTRODUCTION

Sophisticated computer-aided design (CAD) systems have been developed to incorporate solid modeling techniques that can use unambiguous representations in 3-D to represent parts; but this development has largely been independent of the growth of computer-aided manufacturing (CAM), which was a result of the introduction of numerical control in manufacturing machines. Computer-aided process planning (CAPP) attempts to achieve automatic meaningful communication (currently still dependent on human intervention [2]) between these “islands of automation”, in order to reduce lead times and production costs. To achieve this goal, manufacturing feature recognition from the CAD model is necessary.

Most CAD systems on the market support solid modeling techniques that are hybrids of the following approaches: constructive solid geometry (CSG), feature-based representations, and sweep operations. While feature-based modeling readily yields manufacturing features, it is possible to query CAD systems by using their application programming interface (API), for the information provided by the designer while taking any of the other approaches. However, most feature recognition systems tend to ignore this information, and depend solely on the B-Rep of the final solid model, or 2D orthographic representations. This paper proposes a method that exploits the information contained in sweep operations (more specifically, extrude operations) to recognize manufacturing features in the part.

The organization of the paper is as follows. Section 2 covers the previous work that has been done in the area of feature recognition. Section 3 describes the structure of the feature recognizer, while Section 4 elaborates on the

* Corresponding author: Tel.: + 1-740-593-1545; Fax: + 1-740-593-0778; E-mail: sormaz@ohio.edu
Flexible Automation and Intelligent Manufacturing, FAIM2009, Middlesbrough, United Kingdom

algorithms and methodology followed to achieve the desired results, documented in Section 5. The primary CAD system used in this analysis is Unigraphics NX from EDS.

2. APPROACHES TO FEATURE RECOGNITION

Since Kyprianou’s seminal work [3] in 1980, this field has been extensively researched, and several approaches have been developed to decompose the solid model of a part into features that may be understood from a manufacturing point of view. Based on the data used by these approaches to extract manufacturing features, they may be broadly classified into three types: 3D B-Rep based methods, 2D based methods and 3D CSG based methods. While most of these approaches are based on geometric reasoning, artificial intelligence techniques such as rule-based systems [3] and neural networks [4] have also been used to extract features. This section discusses some of these approaches.

2.1 B-REP BASED APPROACHES IN 3-D

Paper [5] documents previous work done in feature recognition in the manufacturing domain, until the year 2000. According to the authors, three approaches seem to be the most active out of the large amount of available literature: graph-based, volumetric decomposition based, and hint-based approaches. These approaches worked on the B-Rep model, which is the dominant representation scheme for solid models in most CAD systems.

Graph-based methods take as input an Attributed Adjacency Graph (AAG), which expresses the relationships between the faces and edges of a B-Rep model. For instance, an AAG might have faces for nodes and edges for arcs joining these nodes, with additional information such as concavity/convexity attributed to the arcs. The algorithms then divide this AAG into sub-graphs and search them for predetermined patterns representing feature profiles [6]. These methods, while applicable to recognizing isolated features, face increasing complexity and difficulty in recognizing interacting or nested features. Some of these problems were overcome by subsequent work through the introduction of missing arcs (or face-edge relationships lost because of intersecting features), but a complete solution could not be guaranteed.

The recent approaches are hint-based methods, which look for traces of features (hints) in the B-Rep of the part followed by feature completion procedures that check for various criteria to ensure the validity of the recognized features [8]. They are base on the reasoning that a feature is represented as the removable volume by a single machining operation, and only those features need to be investigated which leave a trace in the part body. For example, the trace for a pocket would be its floor, or that for a hole would be a cylindrical surface. These traces are then ranked and analyzed for completion. If a valid feature results, then the delta-volume and the hint sets are updated to avoid redundant recognition. These methods have the advantage of being able to handle isolated, interacting, and nested features using the same reasoning.

2.2 APPROACHES IN 2-D

These approaches focus on the vast amount of engineering design data recorded in 2D drawings and CAD systems before the 3D CAD became popular, and attempt to extract manufacturing features from the orthographic views [11] of a given part. Both isolated and intersecting features have been recognized [12]. Reference [14] attempts to integrate this approach with Automatic Visual Inspection (AVI) systems, reasoning that the input to both systems is identical when hidden lines are removed from the CAD drawings. The information loss of hidden line removal is made up by using the isometric projection of the part. Reference [15] applies hint-based reasoning in this domain. This work also provides a summary of previous 2D feature recognition work.

Feature recognition in the 2D domain is less computationally intensive than its 3D counterpart because of the reduction by one dimension in the data structures required to represent the part in formation. However, there has been a distinct domain separation between 3D and 2D methods in terms of the input, with [12] even converting 3D solid models to 2D drawings before they are input to the FlexiCAD system. The method proposed in this paper analyses 2D in formation that is integral to the 3D design of the part. In that sense, it attempts to synthesize the simplicity and consistent reasoning of the hint-based approach with the speed of working in 2D.

3. METHODOLOGY

This section describes the methodology for recognition of manufacturing features from extruded (or sweep) operations. A sweep operation is a term used to refer to the region defined by the movement of an object along a
specified path. Solid models may be created by sweeping a curve, surface or solid (called the generator) along the path (trajectory). Uses of such solid models include a variety of applications, ranging from collision detection in mechanism design to process visualization in manufacturing. In the case of solid modeling, sweep operations may be used to create mechanical parts with constant cross-sections [18], by sweeping a set of curves or a surface along another set of curves.

This section explains the steps in the feature recognition procedure, namely, hint generation, hint verification in 2D and 3D, and, in the end, hint validation and feature generation.

3.1 HINT GENERATION

The first step in the process of feature recognition is the generation of feature hints from the nominal geometry. A hint is a pattern in the part geometry which indicates that it was produced by a certain type of feature, which in turn corresponds to a specific machining operation [8]. Hints are usually implemented as incomplete instances of features, with some parameters being supplied by the observed patterns.

The focus of this thesis is hint generation from the profile data of the extrude operations. In other words, traces of machining features are sought in the curves that make up the profile of each extruded operation.

Three main criteria are used to generate feature hints from the extruded operation and its profile:

Presence. The necessary condition to check for presence is whether the extruded operation still has at least one face associated with it. This can be verified from the BRep information of the part. If no faces make up the extrude operation, it has been subsumed by a larger operation; hence it is not considered for hint generation.

Feature sign. This information is used in conjunction with the loop information to determine which patterns of curves may result in material removal from the final part.

Nature of the loops. Depending on the feature sign of the operation and whether the hint is on the outer or inner loop, it can be inferred whether material is added or removed.

3.1.1 Finding the Material Side

The first step in hint generation is determination of the material side for each sweep profile loop. The material side of an extruded operation is inferred using the following criteria:

- In an operation that carries a feature sign create, unite or intersect\(^\dagger\): – An outer loop has material on the inside, i.e. in the opposite direction of the loop edge normals. – An inner loop has material on the outside, i.e. in the same direction as the loop edge normals.

- Operations with the feature sign subtract behave in the opposite way to above, i.e. an outer loop has material outside it (same as the loop normal direction), and an inner loop has material inside (opposite to the loop normal direction).

Material side behavior is determined differently for outer and inner loops of the profile based on loop normal direction. CAD systems used were reported to be independent of loop types and feature sign; therefore preliminary analysis was performed to compare loop edge normals and understand material behavior. Based on material behavior, loops are divided into two categories: a) CI/SO (Create Inner / Subtract Outer) The loops in this category are either inner loops of extruded operations having feature sign create, unite or intersect, or outer loops of operations with feature sign subtract. For these loops material is always outside them; and b) CO/SI (Create Outer / Subtract Inner) The loops in this category are either outer loops of extruded operations having feature sign create, unite or intersect, or inner loops of operations with feature sign subtract. For these loops material is always inside them.

3.1.2 Hole Hint Generation

The presence of a hole is suggested by a concave cylindrical surface in the part BRep [10]. In an extrude operation, such a surface may be produced by an arc such that its included angle is greater than \(\pi\) and the material side is on its outside. The tapered end of a drill is disregarded in this thesis for the sake of simplicity. The included angle criterion is imposed to ensure machinability, since it is destructive to machine a hole which is not adequately

\(^\dagger\) Operations create, unite, intersect and subtract refer to Boolean operations as defined in Unigraphics CAD modeling system [17]
supported by material. The convexity of the arc is also determined in the following terms: an arc is defined as concave if its curvature is such that it cuts into the polygon formed by the loop, and as convex if it projects outward.

3.1.3 Slot Hint Generation

A slot is a machining feature produced by a linear sweep of a milling cutter. The surest hint of a slotting operation in a part is the presence of parallel faces that face each other and overlap each other when viewed in a direction normal to either of them [8]. These conditions indicate that they may have been produced by a cutting tool in a single operation. When analyzing the profile of an extrude operation, this pattern corresponds to a pair of parallel lines such that:

- The lines “overlap”, i.e. when either line is projected on the other, there is a common region occupied by both the first line and the projection of the second, i.e. \( l_1 \cap p(l_2) \neq \emptyset \).
- The lines are situated so that the material side is on the opposite side of the line when compared to the other line, ensuring that the faces produced are “facing” each other.

Application of those two criteria lead to development of a decision tree (see Figure 2) which, starting with two parallel lines (on one or two loops) shows necessary tests to be performed in order to establish validity of a slot hint. As Figure 2 show it includes rules for loop type, normals of lines and gives a validity value of a hint.

![Figure 1. Examples of valid slot hints: Feature sign subtract (c) Both inner, (d) Both outer and (e) Inner-outer. Arrows indicate loop normals.](image)

![Figure 2. Decision tree for slot hints](image)

3.2 Hint Verification

The generated hints must be verified through a process of completion and validation in order to yield valuable manufacturing feature information. This process happens in two phases, viz. 2D verification and 3D validation. Each phase is explained here in greater detail.

3.2.1 2D Verification

This phase consists of further 2D geometric reasoning to determine the maximum extent of the feature hint in the profile before it intrudes with other parts of the profile.

3.2.2 Hole Hint

The arc that defines the hole hint is completed to a full circle. This circle is then compared with other curves from the profile to check for intersections and enclosed curves. If it intersects the profile, or there are enclosed curves, it means that the hole feature would intrude into the part and that the hole is not valid. If neither of these cases is encountered, the hint is ready for 3D verification.

3.2.3 Slot Hint

Reasoning is performed in the 2D phase to find the limits of the slot by finding the maximum 2D length the slot can have without interfering with the rest of the profile. This is done by taking the following steps. First, the lines representing the slot walls are trimmed to the region of overlap. As mentioned earlier, the overlap refers to the common region between one line and the projection of the other line on it. The starting points and ending points of the trimmed walls are joined to form two lines called the “top” and “bottom” respectively (see Figure 3). These are
the default limits of the slot hint, which need to be replaced, if possible, by other lines that define the length of the hint more accurately.

![Figure 3](image1.png)

**Figure 3.** Creation of default slot hint.

The default slot hint represents the minimum area in 2D that the slot occupies, based on what is known from the walls that give rise to the hint. The next objective is to find what kind of limits the rest of the profile places on the area of this hint. To do this, all the curves of the profile are searched to find every curve that is: a) oriented so that material is on the appropriate side of it in order to enclose the slot, and b) situated so that some part of it “overlaps” the width of the slot. The procedure is illustrated in Figure 4.

![Figure 4](image2.png)

**Figure 4.** 2D verification process of default slot hint.

This is done in an effort to find the curves in the profile that may directly affect the length of the slot hint. Also, depending on whether any of these curves are present in the interior of the rectangular shape defined by the initial slot hint, the hint may need to be partitioned into several smaller child slot hints. To achieve this reasoning, each of these curves is processed in the following way (refer to Figure 4):

- Each curve is “clipped” between the walls of the slot hint.
• The clipped curves are separated into two sets, depending on which side they enclose the slot. If they enclose it in the same direction as “top”, they are called top-candidate curves, otherwise they are known as bottom-candidate curves; If \( N_c \) represents the normal of line \( c \), and \( N_a.N_b \) is dot product of two vectors then these rules apply:
  - If \( N_c.N_{top} > 0 \) line \( l \) is a top-candidate.
  - If \( N_c.N_{top} < 0 \) line \( l \) is a bottom-candidate.
  - If \( N_c.N_{top} = 0 \) line \( l \) is ignored.

• For each top-candidate curve, the point closest to the “bottom” is computed, and for each bottom-candidate curve the point closest to the “top” is computed. These points represent the limit of tool travel in that direction, if that particular curve were chosen as the slot’s limit.

• These top and bottom candidates are clustered into pairs so that they enclose an “empty” section of the slot hint.

• Each of these pairs, along with the walls of the initial slot hint creates a child slot hint.

• Each of these children is further examined for interferences with other curves in the profile. Valid ones are passed on to the 3D phase.

Thus, each initial slot hint now consists of one or more child hints, each of which is a rectangular shape that defines the limits of the slotting operation in that dimension.

The execution of this algorithm on one slot hint of a test part is shown in Figure 5. Extruded model is shown in part (a) of the figure. The initial default slot hint is shown in part (b) of the figure. The top and bottom candidates generated are shown in blue in parts (c) and (d) respectively. Part (e) shows the candidates together and identifies possible pairings, and (f) shows the resulting child slot hint. In parts (b) through (f), the wall curves are shown in red, the top and bottom curves (the default top and bottom in case of parts (b)-(e), and the revised top and bottom for the child hint in part (f)) are shown in black, and the candidate curves are shown in blue. Notice how the candidate curves have been “clipped” to the region of influence.

3.2.4 Defining the Slot’s Orientation.

Even after the 2D extent of the slot hint is found, the actual direction in which it was milled is yet to be decided. In other words, it is not known at this stage whether the rectangular shape that defines each slot hint is the cross-section of the slot or its footprint. To determine the slot normal direction, the slot hints are divided into closed and open slots. If both the “top” and “bottom” of the slot hint are defined by profile curves, then it is regarded as a closed slot. In this case, the slot normal is perpendicular to the plane of the profile. If the slot hint is not closed on any one side by a profile curve, i.e. it retains its default “top” or “bottom”, it is regarded as open. In this case, the slot normal is parallel to the direction of the lines that form the walls of the slot hint. Once the slot normal direction has been determined, the slot hint is ready for 3D verification.

3.3 Verification in 3-D

This phase begins by creating a feature solid from the 2D-validated hint. The 2D verification phase provides the “footprint” of the feature, while the length of the extruded operation provides the maximum extent of impact of this feature hint on the final part volume. Hence, the first and most important step in hint completion is to apply the limits of the extruded operation to the new profile generated by the identification of a feature hint. In case of a hole hint, a cylinder is created with the same radius as that of the arc that produced the hole hint, with the corresponding center, and length equal to the length of extrusion. For each slot hint, a block is created with its length and width equal to those of the slot hint, and height equal to the length of the extruded operation. In case of pocket hints, the profile and extrude depth are used to create an extruded operation. The locations of these feature solids are found by using the inverse of the transformation that was used to convert the 3D extruded profile curves to 2D. Along with the creation of the solids, their parameters are used to instantiate machining feature instances of the classes hole, slot or pocket.

The completed feature bodies are used to validate the feature hints. This is done by checking the respective feature volume and part volume for interference (i.e. a non-null regularized intersection between the part and feature volumes: \( V_f \cap V_p \neq \phi \)). If there is an interference between the volumes, this means that the designer has used at least one subsequent extruded operation to deliberately create material directly in the path of the proposed feature; and so the hint is rejected, and its machining feature instance is discarded. If the volumes are touching but there is no interference, the feature is accepted and stored in the feature model.
4. **FEATURE RECOGNIZER PROTOTYPE**

This feature recognizer has been implemented as part of the Intelligent Manufacturing Planner (IMPlanner) prototype system, currently in development at Ohio University [19]. The feature recognizer serves to increase the versatility of the system by enabling it to handle part models that have been designed using hybrid techniques with both design features and extruded operations.

The feature recognizer prototype consists of three functional modules (Figure 6): the Geometry Query Module (GQM), the Geometric Tool Kit (GTK), and the Feature Identification Module (FIM), which consists of Feature Hint Generator, Feature Completer, and Feature Validator.

These modules interact with the CAD system (Unigraphics NX) to extract the relevant part information, process it, and generate a machining feature model. Each of these modules is briefly explained in the following sections. Section 5.3 describes the details of implementation and execution steps of these modules.

5. **CASE STUDIES**

Some of the parts used as examples for verification of the described methodology are taken from the National Design Repository [20]. The examples presented in this section are introduced with its part geometry, giving the reader an insight into how the geometry was created using extruded operations. Each part is taken through the various stages in the process of feature recognition and only the final results are presented, illustrating the performance of the procedure. The logic and accuracy of the feature recognition process is examined in each case and conclusions drawn about the applicability of the process are given in [21].

5.1 **SLOT TEST**

The design of this part was developed at Ohio University to test the algorithms of slot hint recognition. This comparatively simple part contains only two extruded operations: the first one has feature sign `create` and the second has feature sign `subtract`. The CSG tree with these operations is shown in Figure 7a. The profiles of both the operations contain rectilinear shapes, resulting in 14 possible slot hints. Of these, 8 hints were identified as invalid and 6 hints were validated and corresponding 6 slots created as shown in Figure 7b. The system does generate duplicate hints, but these have been removed for clarity.

![Figure 7. Slot Test example: a) Solid with extruded operations, b) Valid hints generated and the features they represent.](image-url)
5.2 COUPLING

This part is a simplified version of a design taken from the NIST design repository [20]. Again, this part contains only two extruded operations as shown in Figure 8, but it is different from the previous part in that the second extruded operation has feature sign intersect. Total of 6 slot hints generated, three were discarded and three slots were created from valid hints. A hole feature was also recognized in this example.

The most obvious thing to note in this feature model is the fact that the steps on the “outside” of the part have not been recognized. This is because the algorithms developed at the time of writing this paper did not include analysis of step features, i.e., slots are only recognized if there is a guarantee of the presence of parallel faces in the part. The only place these occur in the part are in the center, where the algorithm performs satisfactorily to produce the required features. The accurate recognition of the hole feature vindicates the assumption that extruded operations with feature sign intersect behave in a similar fashion to those with feature sign create and unite as far as their profiles are concerned.

![Figure 8. Coupling example: a) Solid model with extruded operations, b) Two slots and a hole recognized](image)

6. CONCLUSIONS

The research presented in this paper has extended knowledge in feature recognition with contributions in the following areas:

- Development of feature recognition procedure for solid models designed using extruded operations. Extruded operations are very flexible, powerful, and they are also used in a variety of mechanical design applications.

- Reduction of hint generation (and verification) to a 2D problem. While simple rotational parts and swept parts have been analyzed from a 2D perspective before, most of these approaches are syntactic and involve little or no geometric reasoning to obtain manufacturing features. The present approach applies hint based reasoning similar to B-Rep based approaches, but reduces such reasoning to 2D profiles instead of the entire 3D part. This approach makes the calculations simpler since calculations in 2D are easier than in 3D. The approach is also more efficient because many invalid hints may be rejected in 2D instead of in 3D.

- Integration with existing process planning software. The machining feature recognition prototype has been integrated with IMPlanner, and its results are incorporated into a feature model that can be used for downstream process planning activities,
• Direct use of CAD model while maintaining system independence. The feature recognition system itself operates on facts and objects replicated from CAD model input, so that while the system can directly use CAD files, its algorithms and operation are independent of the CAD system itself.

The current implementation does not have a mechanism to eliminate from reasoning those curves that do not create part faces. Currently there is no notion of stock when analyzing profiles for feature hints, and procedure needs to be introduced for more robust reasoning. This leads the algorithms to miss some curves that produce part faces instead of stock faces.

7. REFERENCES

Design of Seat Belts: Safety and Comfort for Expectant Mothers

Lily Amelia*, D.A. Wahab, S. Abdullah and K. Khalid

Department of Mechanical and Materials Engineering
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, Malaysia

ABSTRACT

This paper reports a study on the perception of expectant mothers on comfort, ease of use and safety of the existing automotive seat belt. The study was performed through interview and survey on medical doctors and patients in three maternity clinics in the country. Results from the survey showed that the majority of the respondents feel uncomfortable when using seat belts especially at the abdomen and chest parts. 52% of the respondents feel insecure as to whether the seat belt can protect them and their fetus during minor jerking. For that reason, 75% of the respondents agree that the seat belt need to be redesigned for comfort and safety. Based on the findings, design modifications were performed on the existing seat belt, taking into consideration ease of use, comfort and safety. Modifications to the seat belt include improving webbing width and thickness, incorporating a retracting system that provides comfort to the users and an elastic buckler for maintaining the belt position to be below the abdomen. For validation purposes, the maximum stress and deformation of the new seat belt design were analysed and found satisfactory.

1. INTRODUCTION

The number of road accidents in Malaysia tends to increase every year. In 2005, there were 328,368 road accident cases in Malaysia with 6 188 deaths. In 2006, these figures increased by 4 % to 341,232 accidents with 6,287 deaths [1].

Expectant mothers are expected to be among the highly risk people in any car accident. The impact of crashes to an expectant mother may cause severe maternal and fetal injuries such as miscarriage, especially if they do not use a safety seat belt during the driving. It is imperative for a seat belt to be used in order to protect the expectant mothers and their fetus from serious injuries.

Like other users, the expectant mothers are required to wear seat belt while seating in a car. However, they often ignore the regulation due to discomfort from using the seat belt. Some of them feel pain at their chest and abdomen. Seat belts are often too tight and this leads to a stronger pressing to the abdomen. This condition becomes even more severe at the later pregnancy periods.

Many kinds of seat belts have been designed since the lap belt was first introduced in 1930. Volvo introduced 2-point cross chest seat belt in 1956, while 3-point safety belt (Figure 1) was patented by Nils Bohlin in 1958. Since then, several innovation to the seat belt design have been introduced including the addition of automatic and emergency locking retractors as well as pre-tension retractor [2].

Design of seat belt for expectant mothers is important to ensure their safety as well as to reduce discomfort. Nakahara et al. [3] stated that the expectant mothers need to ensure that the seat belt does not twine around their neck and the belt should cross over the shoulder and should not be pulled out under the armpit. The shoulder belt should be placed on the chest while the lap belt is placed below the belly. The lap belt should not be placed at the middle of abdomen to avoid pressure to the fetus during minor jerking. Hitosugi et al. [4] stated that the expectant mothers should keep the distance of their abdomen from the steering. Based on previous study, the safe distance between expectant mother and the steering is around 14.5 ± 5.8 cm on condition the car speed is less than 56 km/h and the expectant mother’s weight is less than 75 kg.

* Corresponding Author: Tel: +6016-9159379, Fax: +603-89259659, E-mail: lilya@eng.ukm.my, dzuraida@eng.ukm.my
This study aims at developing a new design of seat belt for expectant mothers. In order to identify the requirements of the seat belt, survey and interview were conducted. Based on the survey, an existing 3-point seat belt design was improved in order to enhance comfort and safety of the seat belt.

2. METHODS OF STUDY

In order to identify the design requirements of car seat belt for expectant mothers, surveys were conducted in 3 maternity clinics that are located in three districts in Malaysia namely Kajang and Bangi (in Selangor state) and Muar (in Johor state). The number of respondents from these three clinics is 64 people that consist of patients and female doctors as presented in Table 1. The respondents were selected based on the criteria that they are pregnant or have been pregnant and have used a car during their pregnancy periods. From the survey, 98% of the respondents use car as their main transportation vehicle. 5% of the respondents have experienced minor jerking during their pregnancy.

Table 1: Number of respondents

<table>
<thead>
<tr>
<th>No.</th>
<th>Maternity clinic</th>
<th>Number of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Bandar Baru Bangi maternity clinic</td>
<td>21</td>
</tr>
<tr>
<td>2.</td>
<td>Kajang maternity clinic</td>
<td>13</td>
</tr>
<tr>
<td>3.</td>
<td>Muar maternity clinic</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>64</td>
</tr>
</tbody>
</table>

The questionnaire consists of twenty questions and asks the respondents about their priority when using seat belt, their body sizes, their frequencies in driving during pregnancy, feeling of comfort or pain in wearing seat belt, their opinion in whether wearing seat belt can protect them and their fetal, views on the existing design of seat belt and their expectations in the design of seat belt for expectant mothers. For every question, the respondents are asked to specify their level of arguments based on five Likert scales as follow: 1. strongly disagree, 2. disagree, 3. uncertain, 4. agree and 5. strongly agree. Based on these surveys, some design concepts of seat belts for expectant mothers are generated. The Pugh Concept Selection method was applied to select the best design concept. Modifications to the existing seat belt are performed by taking into consideration safety and comfort. Finally, the new design was validated by analysing stress and deformation using finite element method.

3. RESULTS AND DISCUSSION

3.1. SURVEY RESULTS

The survey shows interesting results in relation to the use of seat belt among expectant mothers and their opinion about the design of seat belts. Awareness on the importance of wearing seat belt was found to be very high. 95% of the respondents stated that wearing seat belt has been their main priority while 5% of the respondents is uncertain and disagree with this statement (see Table 2). Most of them believe that wearing seat belt can avoid serious injuries or even death during jerking.
Design of Seat Belts: Safety and Comfort for Expectant Mothers

The number of respondents that feel uncomfortable when using seat belt is quite high, i.e. 39 %, while 23 % of the respondents are uncertain. Only 38 % of the respondents feel comfortable with the seat belt. As a consequence, they only use seat belt for particular occasions such as driving in highways. Comfort in wearing seat belt seems related to the size of the expectant mothers. 41 % of the respondents claimed that they are of small built while 59 % are of medium and big size. The body size and the car seat position have strong relationship with comfort when using the seat belt. With a larger body size, the car seat position will be too upfront, hence the webbing will be more retracted. As a consequence, the seat belt will impose a higher reverse force that causes the webbing to bind the users tightly, causing them to feel uncomfortable.

<table>
<thead>
<tr>
<th>No.</th>
<th>Features</th>
<th>The percentage of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Wearing seat belt is the main priority</td>
<td>Very agree</td>
</tr>
<tr>
<td>2.</td>
<td>The respondents do not feel comfortable when using seat belt</td>
<td>65 %</td>
</tr>
<tr>
<td>3.</td>
<td>The respondents do not feel comfortable when sitting on the car seat</td>
<td>8 %</td>
</tr>
<tr>
<td>4.</td>
<td>Respondents feel pain or discomfort at their abdomens when wearing seat belt</td>
<td>5 %</td>
</tr>
<tr>
<td>5.</td>
<td>Respondents feel pain or discomfort at their chests when wearing seat belt</td>
<td>11 %</td>
</tr>
<tr>
<td>6.</td>
<td>Respondents feel confident that their fetus are safe when they are in car</td>
<td>9 %</td>
</tr>
<tr>
<td>7.</td>
<td>Respondents think that the seat belt can protect them and their fetus from injuries</td>
<td>30 %</td>
</tr>
<tr>
<td>8.</td>
<td>Respondents think that the seat belt must be redesigned for comfort and safety</td>
<td>26 %</td>
</tr>
</tbody>
</table>

Only 27 % of the respondents feel uncomfortable when sitting on the car seat. This percentage is less than the percentage of respondents that feel uncomfortable when using seat belt which is 39 %. This indicates that the main factor that causes discomfort to the expectant mothers is wearing the seat belt rather than the car seat. In relation to this requirement, this study will focus on the design improvement of the seat belt.

The survey indicated that 49 % of the respondents feel pain or discomfort at their abdomens when wearing the seat belts as compared to 28 % of the respondents that do not feel pain or discomfort and 23 % is uncertain. This may be caused by the webbing that binds the abdomen. Other cause is the webbing position may be in the middle of the abdomen and not below the abdomen so as it depresses the abdomen and causes pain. Meanwhile, many respondents also feel pain or discomfort at their chests when wearing the seat belt. The survey shows 41 % of the respondents feel pain or discomfort at their chests as compared to 34 % of the respondents who do not feel pain or discomfort and 25 % is uncertain. When the webbing crosses the chest, it may impose a high force to the chest and causes pain. This is influenced by the body size of the expectant mothers as well as the car seat position.

Most of the respondents are confident that the safety system inside the car such as seat belt and air bag system will protect them and their fetus. 68 % of the respondents feel confident that their fetus will be safe when traveling in their cars. Only 5 % of the respondents do not feel confident and 27 % is uncertain.

Almost half of the respondents (48 %) are confident that the seat belt will prevent them from miscarriage or injuries if they are involved in minor jerking. However, quite many expectant mothers feel unconfident with the safety provided by the seat belt. 15 % of the respondents feel that even though they wear seat belt they may still get injuries during minor jerking and 37 % is uncertain. Therefore, besides comfort, the aspect of safety must be included in the new design. In consequence, the majority of respondents (75 %) agree that the seat belt has to be redesigned for their comfort and safety. Only 8 % of the respondents disagree and 17 % is uncertain. Besides seat belt, the redesign of car seat is also required by the expectant mothers. 60 % of the respondents want the existing seat belt to be redesigned, while only 10 % of the respondents think that the existing design is comfortable and safe for them.

As a conclusion, the survey indicated that expectant mothers require a safe and comfortable seat belt to protect them and their fetus. Some requirements that need to be considered in designing the seat belt include the seat belt does not cause pain and discomfort, can avoid serious injuries and miscarriages, applicable for any type of car, ease of use, less number of webbing crossing the body, protect the abdomen, low maintenance cost, well-functioned...
during minor jerking, ease for body movement, ease for parts replacement, webbing do not break easily and less displacement during minor jerking.

3.2. SEAT BELT REDESIGN

Modification to the existing seat belt is performed by taking into account the above requirements. In order to avoid pain when wearing the seat belt, the seat belt design should minimise webbing retraction force that presses the body. Minimisation of pressing force is also considered to avoid injuries and miscarriages during minor jerking besides the body movement space and displacement distance at the abdomen part. The seat belt design is required to minimise the body movement space and the displacement distance to the front as well as to the right and the left. The need for the seat belt to be applicable for any car and ease of use is related to design form, the number and position of buckle button and webbing roller torque. Buckle button should be easy to operate and the position of buckle button should be close to the user. Webbing roller torque should allow ease of use therefore the user will feel more comfortable. The expectant mothers also want the seat belt to protect the abdomen during minor jerking. At that time, the webbing is supposed to be automatically retracted and will bind the body tightly. Elastic webbing is needed to avoid the webbing from binding the abdomen too tight therefore it does not affect the fetus. The webbing is also required to be functional during minor jerks. This is related to a good mechanism of acceleration and de-acceleration sensors that leads to a short reaction time for webbing in protecting the user during minor jerking. Other user’s requirements are the seat belt has low maintenance cost and ease for part replacement. The expectant mothers also require that the webbing do not break easily. This is affected by the strength of material or yield point. The webbing should be made from strong materials to hold high tension during minor jerking. Design sub functions are derived based on expectant mother’s requirements as presented in a morphology chart below (see Figure 2).

![Morphology chart](image)

Figure 2: Morphology chart

For the first sub function, i.e. to avoid webbing from broken, some webbing width and thickness alternatives are proposed and nylon is chosen as the webbing material. For the second sub function, the webbing position is
maintained near the shoulder and far from the neck to avoid serious injuries. To enable this position, three design alternatives are proposed including the use of an adjustable arc column, the use of a chair hook or the webbing move direction is designed from the back of seat.

The third sub function is to maintain the user’s comfort. This requires the webbing retraction force to be less than 1 N. Three design concepts are proposed for this requirement including the use of a torsion spring, an electric motor or using an additional lock clip. The fourth sub function is for the seat belt to maintain the user’s position in the car seat and minimise the displacement during minor jerking. This requires the webbing to bind through the pelvis bone. There are three alternatives to allow the webbing to bind through the pelvis bone such as the buckler is welded to the seat platform, using an elastic buckler or using a buckler that can be adjusted to the left and right side. A pre-tension sensor, centrifugal clamp or an electric motor in retractor enable the webbing to bind tidily. In addition, three alternatives of seat belt are proposed to minimise the side displacement such as 3-point seat belt, crossing belt or the webbing can expand as similar to the air bag system.

Based on this morphology chart, sixteen design concepts are generated by combining every proposed design concept in each sub function. For example, concept 1 proposes webbing width and thickness of 4.7 cm and 0.1 cm respectively, the application of adjustable arc column, torsion spring, buckler that is welded to the seat platform, use of pre-tension measurement and 3-point seat belt type. Meanwhile, concept 2 proposes webbing width and thickness of 4.7 and 0.15 cm respectively, webbing from the back of seat, torsion spring, buckler welded to the seat platform, the use of pre-tension measurement and 3-point seat belt type. Similar approach in combining the different solutions proposed in Figure 2 resulted in other fourteen design concepts. For purposes of this paper, the selected design concept is explained in detail.

The best design concept is selected through two selection stages using the Pugh Concept Selection Matrix [6]. In the first stages, the sixteen design concepts are compared with the datum which is a 3-point seat belt. The comparison is based on twelve criteria that have been identified from the survey. The sign ‘+’ means better than datum, ‘S’ means same with datum while ‘-’ means worse than datum. As overall, concept 1, 6, 9, 13 and 15 reach the highest five ranking, therefore they are selected to the second stage. In the second stage, these five concepts and one additional concept which is the combination of concept 1 and 13 (concept 1+13) are compared. In this stage, concept 1+13 achieves the highest ranking and is selected as the best design concept of seat belt (see Figure 3).

![Figure 3: Pugh concept selection matrix](image)

Concept 1+13 is a modification of 3-point seat belt (Figure 4). In the concept 1+13, an adjustable arc column is applied to enable the webbing position to be near the shoulder and far from the neck. Compared with the original design of 3-point seat belt, the retraction system is modified using a torsion spring as one of the component. Elastic buckler and pre-tension sensor are used to maintain user’s position in the car seat and minimise displacement. Webbing width and thickness are designed as 5 cm and 0.1 cm respectively. Concept 1+13 is better than the original 3-point seat belt and the other proposed concepts in protecting the expectant mothers from serious injuries during minor jerks. The retraction system in concept 1+13 could minimise the body displacement during minor jerking. As a consequence, it gives more protection to the expectant mothers from miscarriages and injuries.
Concept 1+13 consists of six components including retraction system, adjustable arc column, tongue, end bracket, buckler and webbing. In concept 1+13, modifications are performed in two components which are the retraction system and the buckler. There are nine components in retraction system that are redesigned including host, shaft, torsion spring, torsion spring cover, lock, air inducer, rod pinion, roller and screw (Figure 5). Host is made from carbon steel that has higher modulus of elasticity and modulus of rigidity compared with non-corrosive steel and copper. The host dimension is 90.3 mm length, 61 mm width and 186.5 mm height and only suitable for webbing that has width less than 51 mm. Shaft is also made from carbon steel and there is a spur gear in the end of shaft. During minor jerking, the gear will be actuated by the rod pinion to retract the webbing. As a consequence, the webbing will bind the user tidily. After the gear rotates 1.5 cycles, rod pinion will stop and the shaft will be locked. Therefore, the webbing cannot be pulled out and displacement is restrained. Rod pinion is made from carbon steel and has diameter is equal to 22.5 mm and length is 130 mm. During minor jerking, the air inducer will produce a high pressure gas that will push the piston inside the shaft to the upper side. Therefore, this will cause the rod pinion to move upward with the speed 120 m/s and turn the gear. Torsion spring is made from copper sheet roll, while torsion spring cover is made from ABS plastic. In the normal condition, this torsion spring produces the maximum webbing retraction force is 1 N.
cylinder at the end of elastic rod. These semi cylinders function as movement graduation in order to adjust the buckler movement.

![Elastic buckler components](image)

Figure 6: Elastic buckler components

These modifications result in safety and comfort for the expectant mothers. In normal condition, this seat belt design retains the webbing loose so that the users feel free and more comfortable. This is caused by the retracting system that maintains the retraction force of less than 1 N resulted by the torsion spring. If there is a minor jerk, the pre-tension sensor will send information to the processing unit to operate the gas inducer. Gas induction will cause the rod pinion to move upward with the speed of 120 m/s and rotate the shaft gear. Webbing is rolled back to the retractor causing the buckler to be retracted. The elastic rod will bend to allow the webbing to bind the pelvis bone tidily. This mechanism will take place in just 1ms after an accident.

3.3. **DESIGN VALIDATION**

Finite element analysis was conducted to the gear shaft and rod pinion to ensure that the design of these critical components will not fail. By assuming the average weight of expectant mothers is around 75 kg, car speed 56 km/h and maximum deceleration is 50 m/s², the static force to the gear and rod pinion is determined as 4.45 MPa. Using a finite element software, the maximum stress of the gear shaft at this static force is 353.94 MPa, while the tensile stress is 117.98 MPa (see Figure 7). This tensile stress is less than the yield strength of carbon steel which is 207 MPa. The maximum deformation is only 0.04 mm or very small which is in the end of the gear shaft. Therefore, the design of the shaft is found to be safe.

![Maximum and tensile stress of shaft](image)

Figure 7: Maximum and tensile stress of shaft
Figure 8 indicates that the maximum stress of rod pinion is 171.62 MPa. The tensile stress is 133.49 MPa that is also less than the yield strength of the carbon steel. In addition, the maximum deformation is 0.04 mm and can be neglected. Therefore, as a whole the design of rod pinion is found satisfactory.

![Figure 8: The maximum and tensile stress at rod pinion](image)

4. CONCLUSIONS

Expectant mothers require a seat belt that is safe and comfortable. The survey conducted in 3 maternity clinics in central and southern part of Malaysia indicated that the requirements for seat belt for expectant mothers include seat belt that does not cause pain and discomfort, can avoid serious injuries and miscarriages, applicable for any type of car, ease of use, less number of webbing crossing the body, ability to function during minor jerking, ease for body movement, ease for parts replacement, webbing do not break easily, less displacement during minor jerking and low maintenance cost. The survey only focuses on the expectant mother’s requirements on the design of seat belt. Other design requirements for expectant mothers such as car seat and air bag system are not included in this survey. Based on these requirements, the existing design of 3-point seat belt was improved including modifications to the retracting system and the use of elastic buckle. This new design provides safety and comfort to the expectant mothers during normal condition as well during minor jerks. The new design was validated using finite element method and was found satisfactory.

REFERENCES

Ontology-based Data Process for Integrated Data

Xiaomeng Chang¹, Yanfeng Li¹, Brandon Brisbon², Janis Terpenny *¹,³,⁴

¹Departments of Industrial and Systems Engineering, ²Electrical and Computer Engineering, ³Engineering Education, and ⁴Mechanical Engineering
Virginia Tech, Blacksburg, Virginia, 24061, U.S.A.

ABSTRACT

Many companies and organizations make use of commercial software applications in order to enhance the efficiency of design, manufacturing, and sales. Indeed, software applications could support all functions and activities of an enterprise. To be most effective, applications often create or connect to multiple data resources. While much research has been done to integrate and analyze distributed data, significant challenges exist since integrated data is often heterogeneous (of varying types and formats) and comes from different sources. Moreover, there might be redundant or repeated information in the integrated data. In order to deal with these problems, this paper proposes an ontology-based method to help transform the data into a consistent format and unit systems, as well as a method to filter out repeated data. The framework and process of the proposed method are described first. The transformation module that assures data unit consistency is then described followed by a detailed description of the module that analyzes, screens for, and removes repeat data. Finally, the method is demonstrated with a simple example along with the details of the realized system. Ultimately, the advantages of ontology in the representation of relations and reasoning will provide the higher quality data that is so needed for decision support and framework needed by integrated systems.

1. INTRODUCTION

Many companies and organizations make use of multiple commercial software applications in order to enhance the efficiency of design, manufacturing, and sales. Indeed, software applications could support all functions and activities of an enterprise. To be most effective, applications often create or connect to multiple heterogeneous data resources. Much research has been done to integrate and analyze distributed data [1, 2]. Since integrated data comes from multiple heterogeneous data resources, the format of data can be a big problem. Further, data from heterogeneous data resources are likely to be in different units. For example, weight can be expressed in ounces, grams, etc, and could be expressed in mixed unit measures (e.g., 5 pounds and 5 ounces). Similarly, formulas may also be based on different unit systems. Clearly, integrated data needs to be transformed into consistent format to avoid errors.

In this paper, an ontology-based data process method is put forward to transform data into consistent format and unit systems. Ontology is a formal specification of domain knowledge and has been used to define a set of data and their structure for experts to share information in a domain of interest. It is good at the representation and utilization of relations among data. Further, ontology-based tools are efficient in knowledge reasoning. Since there can be abundant mathematical relations among different units in a category, ontology is a good way to represent relations among units and to help reason out formula. For example, in the length category, the ontology can record all the length units and their relations with meter. An ontology-based tool can then be developed to help conclude formulas between any two length units.

Since integrated data comes from multiple heterogeneous data resources, conflict or duplication within the data is a big challenge for its further use. When data comes from different resources, repeated rows or columns occur among data. These repeated data can influence the quality of decision support systems and analysis tools. In recent years, data mining methods, such as clustering, have been used widely to analyze data and group the similar data. This paper also uses these methods to screen out repeated data that exist in integrated data coming from different data resources.

In the sections that follow, the framework and process of ontology-based data process are described first. The unit ontology that supports data transformation is put forward, and the ontology-based transformation module that assures data unit consistency is then described followed by a detailed description of the module that analyzes, screens for, and
removes repeat data. Finally, this method is demonstrated by a simple example and the realized system is presented. Ultimately, utilizing the advantages of ontology in the representation of relations and reasoning, the methods in this paper will provide subsequent decision support tools and systems with data in better quality.

2. Ontology-based Data Process for Integrated Data

In order to deal with the inconsistency of data format and data units as well as to filter out the redundancy in the integrated data, in this section, an ontology-based method to process integrated data is put forward with more details on the unit ontology, ontology-based data transformation and calculation, as well as the data redundancy filtering.

2.1 Framework and Ontology-based Data Process

The framework and process of ontology-based data process for integrated data is shown in Figure 1. There are three modules in the system, including: Unit Identification, Transformation and Calculation, and Data Redundancy Filtering.

![Figure 1: Framework and process of ontology-based data process for integrated data](image)

In previous research, data has been integrated from a variety of resources by a data integration system [1, 2]. After the system receives integrated data from an external data integration system, the units in the integrated data are first identified into different format categories, such as “number operator number unit”, “number unit”, etc. Here the operator is a calculation operator, such as + or -. For example, in the length data, one is “5 feet + 5 inch”. So this data belongs to the “number unit operator number unit” category. After this step, numbers, units and operators in the data can be identified.

In the second step, the unit ontology is queried to get the relations between target unit system (or target unit) and the units in the original data (identified from the Unit Identification model), described in Section 2.2. For example, the length data should be transformed into inch as user required. So the target unit is inch. The relation between inch and all the other length units, such as meter, feet, kilometer, etc., can be retrieved from the unit ontology.

Based on the queried result from the unit ontology, the Transformation and Calculation module (Section 2.3) concludes the unit transformation formula between the identified units (identified from Unit Identification model) and target units. After that, it transforms the original data into uniform data format with the target units based on the formulas. Since the data after the Transformation and Calculation module is originally integrated from multiple separate data resources, there might be redundancy. Hence, the Data Redundancy Filtering module (Section 2.4) filters out the repeated data in order to support the decision support module or system.

2.2 Unit Ontology

Ontologies offer advantages in the representation and utilization of semantics and can be used to represent the relations among units. There have been ontologies to treat units in different systems [3, 4]. In particular, Borchert [4] built an ontology in a style where each relation between units is realized by using a specific program of Java Function Calls. Unit calculation and data transformation can be realized inside Protégé, a popular ontology editor [5]. Because Java Function Calls are integrated to the specific ontology, it is relatively difficult to query out relation information and use the relation outside the ontology model to finish more complex applications. In order to make the system
realize more complex functions, such as complex calculations, and enhance its flexibility, a new unit ontology is built in this paper by using the ontology editor Protégé [5]. The structure of the unit ontology is shown in Figure 2.

![Figure 2: Structure of unit ontology](image)

As shown, there are three classes, including: Category, Unit system and Unit. In the class Category, instances such as length, temperature, etc. can be built. In the class Unit, instances such as cm, mm, inch, etc can be built. In the class Unit system, one instance includes units in different categories with one unit in each category. For example, an instance can includes inch in length, °F in temperature, pound in weight, etc.

In each class there are several properties. For example, the class Category class has three properties: Name, Description (Desc) and Units in the category. The property units in the category can contain several instances in class Unit. And the property Unit is reverse to the property Category of the class Unit. Similarly, the property unit system of the class Unit is reverse to the property Units in the unit system of the class Unit system. Further, in the class Unit, there is a property ‘Relation with standard’ that records the mathematical relation between this unit and the standard of the category. For example, inch and cm are two unit instances whose category is length. In the length category, cm is set as the standard. Then in the inch instance, the value of “relation with standard” is 2.54x, and in the cm instance, the value of “relation with standard” is x. Another example, in the category temperature, there are °C and °F, and °C is set as the standard. Then in °F instance, “relation with standard” is 9/5x+32.

In McGuinness et al.’s KSL wine agent [6], after a meal description is received, a selection of matching wines available on the web is retrieved. This is realized by consulting the ontology, performing queries, and outputting results. It combines a logical reasoner (called JTP [7]), an object-oriented modular reasoning system with an OWL (ontological mark-up language) ontology. This JTP query process was also used in other ontology-based systems, such as [8-11]. Similarly, JTP is used here too to query the unit ontology in OWL.

2.3 TRANSFORMATION AND CALCULATION MODULE

The process of transformation and calculation module is shown in Figure 3. From Figure 3, users can select the target transformation unit or target transformation system. If a target transformation unit is selected, it means the system will transform one specific column of the integrated data that the user selects into the same format with the same unit, which is the target transformation unit. If the target transformation system is selected, all the columns in the integrated data will be transformed and represented in the units of this unit system. The relation between the unit system and units can be queried out from the unit ontology in Section 2.2.
After the target unit(s) is specified, formulas to transform the original unit to the target unit(s) will be formed based on its relation with standard. For example, the target unit for the length unit is meter, and the standard of length in the unit ontology is centimeter. In the unit ontology, meter is equal to 100x, inch is equal to 2.54x, so the formula between inch and the target unit meter is inch=2.54*m/100.

After the formulas are formed the unit in each record is transformed. Table 1 shows some possible cases in this step, where X and Y are number, and X’ and Y’ are numbers and operators according to the formulas in the previous step. For example, in the first case in Table 1, X is equal to 1 and X’ is equal to 1*0.01

Table 1: Cases of unit transformation

<table>
<thead>
<tr>
<th>Before Unit Transformation</th>
<th>After Unit Transformation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [unitA]</td>
<td>X’ [target A]</td>
<td>1 cm to 1*0.01m</td>
</tr>
<tr>
<td>X [unitA] &lt;operator&gt; Y [unitB]</td>
<td>X’ [target A] &lt;operator&gt; Y’ [target A]</td>
<td>1 cm+1 inch to 1<em>0.01m+1</em>0.0254m</td>
</tr>
<tr>
<td>X &lt;operator&gt; Y [unitA]</td>
<td>X’ &lt;operator&gt; Y’ [target A]</td>
<td>1<em>2 cm to 1</em>2*0.01m</td>
</tr>
<tr>
<td>&lt;operator&gt;{ X [unitA]}</td>
<td>&lt;operator&gt;{ X’ [target A]}</td>
<td>√100cm² to √m²</td>
</tr>
<tr>
<td>{ X [unitA]} &lt;operator&gt;</td>
<td>{ X’ [target A]} &lt;operator&gt;</td>
<td>(1cm)² to (0.01m)²</td>
</tr>
</tbody>
</table>

After the unit transformation, all operators will be processed in the calculation step. For example 0.01m+0.0254m will get the result 0.0354 m. Finally, the system can pack the transformation and calculation result, and output it to the data redundancy filtering module.

2.4 DATA REDUNDANCY FILTERING MODULE

One method of identifying repeated data is to calculate the dissimilarity of each data. As shown in formula (1), repeated data is located by identifying those data with difference distance that is zero or close to zero.

\[ d_{Euclidean}(A, B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \]  

(1)

where A and B are the two data to be compared whose dimension is n, and A_i and B_i are the value on the i^{th} dimension.

Further, a clustering method can also be used in the data redundancy filtering module. Clustering is the classification of objects which is represented by data vectors into groups so that objects from the same cluster are more similar to each other than objects from different clusters. The K-means algorithm can be used to divide the whole object set into K subsets (clusters), where K is predefined accordingly [12]. So the objects in a cluster can be displayed by a representative object, such as a dummy object with the average values of the clusters. In this way, the size of data set can be reduced and the very close objects can be screened out.
3. **Example and the Realized System**

This section demonstrates the method through a simple example first. It then illustrates the realized system. In this example, there are two data, as shown in the second and third column of Table 2.

**Table 2: An example of data processing module**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Before data processing</th>
<th>After the first two steps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>2+3.08 cm</td>
<td>2 inch</td>
</tr>
<tr>
<td>2</td>
<td>2 oz</td>
<td>60-3.3g</td>
</tr>
</tbody>
</table>

In this system, the unit identification module reads the data and classifies $A_1$ and $B_2$ into “number operator number unit” format and identifies $A_2$ and $B_1$ into the “number unit” format. The unit ontology is queried to find the unit standard name and unit category. For example, cm is centimeter and belongs to the unit category of length. After a user indicates the target units of transformation, e.g. length should be transformed into centimeters and weight into grams, the unit ontology provides relations between these units with the standard unit in each category. For example, cm is equal to $x$ and inch is equal to $2.54 \times x$, where $x$ is centimeter, the standard unit of the length category in the ontology. Then the formula between units and the target unit can be concluded, which is 1 inch is equal to 2.54 cm. After the formula is formed, the unit is transformed and the number is calculated by the Transformation and Calculation module, as shown in the last two columns of Table 2. Finally, the Data Redundancy Filtering module uses the clustering method or calculates the distance of $A$ and $B$ by formula (1), and finds these two data are the same and therefore filters one out.

The complete system in this paper is still in development. And the realized part is a Java application, as shown in Figure 4. It is a tool that allows a user to scan the integrated data column by column and convert them to a uniform format. The top three buttons control connections to the data integration result because the integrated data might be exported to some specific format or system, such as Excel, Access, and MySQL. The preference button provides the interface to allow the user select the target unit. Once the transformation and calculation are finished, the user can select to “Export” to print results to a text file or to the actual database itself.

![Figure 4: Realized data unit conversion tool](image)

4. **Conclusions and Future Work**

In this paper, an ontology-based data process for integrated data is put forward. Ontology is good at the representation and utilization of relations among data, and is efficient in knowledge reasoning. Utilizing the advantages of ontology in relation representation and realization, methods presented in this work help to process the integrated data that comes from heterogeneous data resources, transform them into the necessary format, and provides...
the following decision support systems and tools with integrated data in better quality. Based on the processed data, the following systems can work more efficiently and accurately.

At present, the data redundancy filtering module is still in development. In the future, this module will be enriched with more algorithms and these algorithms will be compared. The most efficient algorithms will be used in the realized system. In addition, the unit ontology will also be improved with more units and more information.

ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation through Grant No.s EEC-0632758 and EEC-0542084. Any opinions, findings, and conclusions or recommendations presented in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


New Product Development in a Manufacturing Company – A Challenge for Supply Chain Management

Per Hilletofth¹, Dag Ericsson², Olli-Pekka Hilmola³ and Philip Hedenstierna¹

¹ Logistics Research Centre
University of Skövde
541 28 Skövde, Sweden

² School of Engineering
University College of Borås
501 90 Borås, Sweden

³ Lappeenranta University of Technology
451 00 Kouvola, Finland

ABSTRACT

The last decades a new type of business environments has evolved characterized by rapid and volatile demand changes, short product life cycles, and high levels of customized products. The competitiveness of a business in these environments is mostly determined by its responsiveness. This is characterized by the ability to quickly scale up or down the production volume, the presence of an innovative and fast product development, and the quick incorporation of customer requirements into the product development. This paper employs a descriptive single case study approach to illustrate how product development is structured and executed in an international manufacturing company, seeking to realize an innovative, predictable, and efficient product development. The objective is to increase the understanding of how product development and product life-cycles are connected to Supply Chain Management (SCM). Case study findings reveal that the case company after implementing a strategic and structured Product Creation Process (PCP) has improved the efficiency and effectiveness of product development. Findings also reveal that the case company has not yet developed any linkages between product development and SCM. Still, the case company has become aware of this issue due to problems associated with the lack of integration between product development and SCM.

I. INTRODUCTION

Supply Chain Management (SCM) can be defined as a set of approaches utilized to efficiently integrate and coordinate the materials, information and financial flows across the Supply Chain (SC), so that merchandise is supplied, produced and distributed at the right quantities, to the right locations, and at the right time, in the most cost-efficient way, while satisfying customer requirements [1]. One business process that truly affects SCM is the product development process. Without new products, market acceptance and value added packages, an efficient SC is useless, supplying, producing and distributing the wrong products efficiently [2]. However, SCM also affects product development since product competitiveness is not solely restricted to product attributes, but also concerns customer service, and these issues needs to be addressed in parallel [3]. This is particularly the case for consumer markets, where it is important to evaluate and provide accurate procurement and logistics alternatives.

In recent years a new type of business environments – characterized by rapid and volatile demand changes, short product life cycles, and a high degree of customized products – has evolved [4]. In this environment innovative and rapid product development is extremely important, implying that product development must focus on customer needs rather than on technology improvements alone. This means that successful product development requires a deep understanding of customers and their needs. Many companies wish or claim to be customer oriented, however, customer-driven product development is a demanding and difficult task. The voice of the customer must be taken into account in all the phases of product development, both in definition and design phases.

Shortening of product life cycles, more rapid product obsolescence, and the increasing intensity of global competition have driven firms to strive for a rapid introduction of new products [5]. The ability to reduce cycle time in product development is increasingly viewed as a key to innovation success and profitability [6]. First-to-market products may command higher initial prices and gain a dominant market share, as well as customer loyalty [7]. Compressing the new product development process lead times may also result in significant cost reductions. However, shorter product life cycles also implies that product development must be integrated with SCM in order to
successfully introduce products on the market and to ensure that the product assortment is updated according to product life cycles and that obsolete products are properly out-phased [8]. This implies that an effective implementation of the product development process not only enables management to coordinate the flow of new products efficiently, but also to assist in the ramp-up of manufacturing, logistics, marketing and other related activities, that support the commercialization of the product. For this reason, companies need to address all these issues in parallel to be successful.

In this research a descriptive single case study approach is used to illustrate how product development is structured and executed in an international manufacturing company that seeks to realize innovative, predictable, and efficient product development. The aim of our research is to form an understanding of how product development and product life-cycles are connected to SCM. Instead of using an SCM perspective, we outline how the product development is structured and executed in the case company. Our research is mostly descriptive, and aims to shed more light on how SCM and a formalized product development can be connected. The research question is: “What effect will new requirements of a product development process have on SCM?” The primary research approach consists of an embedded single case study, which was considered appropriate in order to gather in-depth data. The case company is Husqvarna, a Swedish manufacturer operating on an international basis in the outdoor power products industry. The remaining of this paper is structured as follows: First we present a literature review of the product development process in Section 2. Thereafter, Section 3 presents and discusses research approach and data collection. Section 4 presents case study findings. In the final Section 5 research is discussed and concluded, and further research ventures are proposed.

2. LITERATURE REVIEW

SCM is recognized as the integration of key business processes across the SC. One of these important business processes is product development and commercialization. The Product Development and Commercialization (PDC) process provides the structure for developing and bringing new products to the market, at best jointly with customers and suppliers [8]. Effective implementation of the process not only enables management to coordinate the flow of new products efficiently across the SC, but also supports the ramp-up of manufacturing, logistics, marketing and other activities related to the commercialization of the product.

Developing products rapidly and moving them into the marketplace efficiently is important for long-term corporate success [9]. In many markets, 40 percent or more of revenues come from products introduced in the prior year [10]. The ability to reduce the time-to-market is key to innovation success and profitability [7] as well as being the most critical objective of the process [11]. As product life cycles shorten, the right products must be developed and successfully launched in ever shorter time frames, so that competitiveness [12] and marketplace differentiation may be achieved. One branch of product development research consists of analytic models to determine the trade-offs between overall product quality and time-to-market [see e.g. 13]. Another branch has focused on identifying specific mechanisms, whereby a firm can improve its product development process [see e.g. 14].

While a considerable body of research has focused on the time-to-market and process improvement problems in isolation, consideration of both design and production (or logistics) decisions to support bringing new products to market has remained relatively unexplored [15]. In their review of product development literature, Krishnan and Ulrich [16] comment that the literature addressing production ramp-up and product-design decision-making is relatively sparse, though notable exceptions exist [see e.g. 17].

Numerous researchers have proposed product development models with 5 to 10 phases for product development [see e.g. 18]. In their review of the product development literature, Karkkainen et al. [19] argue that the product development process consists of three interconnected phases: (1) strategic planning, (2) customer need assessment, and (3) product development. The objective with strategic planning is to ensure clear identification and prioritization of different product development areas by setting specific goals for product development. The goals define the starting point for customer need assessment and product development. Strategic planning builds on current business strategies (e.g. product innovation strategy and brand and design strategy), and serves to clarify business goals before initiating the customer need assessment and the product development phases to avoid the danger of performing the wrong activities. All the people involved in the development activities must understand the goals clearly and keep them in mind during the whole development process. The objective of customer need assessment is to clarify customers’ needs as well as the competitive situation for the company. Customer-driven product development requires proper customer need assessment, which is a systematic activity of gathering and clarifying customer needs, determining product characteristics based on the clarified needs and ensuring that all important needs will be fulfilled. Careful, systematic need assessment helps to focus development efforts and to reduce the need for future design
3. RESEARCH METHODOLOGY

In this paper it was considered that an inductive approach was the most appropriate research strategy since the aim was to enhance current knowledge of how SCM and a formalized new product development could be connected by investigating one international manufacturing company in a qualitative manner. Moreover, due to the context-bound nature of the studied phenomenon, case studies would be an appropriate method. Thus, this paper employs a descriptive embedded single case study [22]. The case company (Husqvarna) is a Swedish manufacturer operating on an international basis in the outdoor power products industry.

One advantage with case studies is the possibility to combine several data collection techniques; in this research empirical data was collected from various sources to enhance understanding by examining the research object from several perspectives. Firstly, this study is based on data gained from in-depth interview with the chief of administration in the product development department; the interview was conducted in 2007. Digital recording in combination with note taking was the main interview method, and the length of interview was 120 minutes. In order to find relevant information the interview was prepared carefully. Furthermore, interviewee was able to read the transcribed interview text afterwards to avoid misunderstandings. Moreover, this study is based on a number of internal (i.e. annual reports and technical reports) and external documents (i.e. industry reports). These were included to provide information to both the industry and the case company’s background.

The data collection has been well documented and digitally recorded, which increases the reliability of the case study. However, it should be noted that all case studies are unique and the companies are continuously changing, meaning that the conditions can never be identical. Two tactics have been applied to increase the validity of this study. Firstly, multiple sources of evidence have been used, and secondly, the draft case study reports have been reviewed by the respondents. Different sources have been used to answer the same questions and therefore triangulation can be said to have been used. The use of triangulation has contributed to improving the rigor, depth, and breadth of the results, which can be compared to validation [22]. However, it also enhances the investigator’s ability to achieve a more complete understanding of the studied phenomenon. Nevertheless, in retrospect, the overall reliability and validity of the study could have been further improved by increasing the number of informants and extending the period of data gathering to encompass multi-points in time rather than providing a retrospective snapshot.

4. CASE STUDY: HUSQVARNA

Husqvarna is the global leader in outdoor power products for forestry and lawn and garden maintenance, as well as cutting equipment and diamond tools for the construction and stone industries. The company sells products to consumers and professionals in more than 100 countries. The largest markets are in North America (54 %) and Europe (40 %), only smaller quantizes are sold in the rest of the world, with the Asian market becoming increasingly important due to the development in China. In 2007, Husqvarna had sales of SEK 33 billion and approximately 16,000 employees.

The case company’s operation is currently organized in two business areas: (1) Consumer products and (2) Professional products. The consumer products share of total sales (67 % of total sales) has increased in recent years; however, the case company makes more profit in the professional sector (33 % of total sales). The products are manufactured around the world and sold under several brands; however, the major share of products is sold under the Husqvarna brand (40%). Some brands are used in both the professional and consumer sector while others are solely professional or consumer brands. The case company’s manufacturing strategy is to manufacture as close as possible to major final markets. Accordingly, most of the products are currently manufactured in North America and Europe. However, opportunities to manufacture components in low-cost countries in Asia and transport these to assembly plants near final markets are being considered. The major issues of this approach are the extensive transportation lead-times (5-6 weeks by sea, air is not a long-term option due to costs incurred) and the risk of a decrease in product
quality. Additionally, the current market in Asia is small (1-3%), making the location of factories there a less interesting option. The case company uses different distribution channels to a certain extent. For professional products it exclusive uses specialized retailers while it for consumer products primarily uses ordinary retailers.

On the consumer market the case company has noticed that it has become increasingly important to continuously develop new or modify old products in accordance with customer preferences. Consumers are more impatient than professional customers who prefer proven functionality and technology. Additionally, cost has become the major order winner while quality is a qualifier. This means that the SC operations need to be cost-efficient and provide high quality output. Moreover, it is very important to have a strong brand that signals high quality and that communicates the company’s values. In order to survive in this environment the case company believes that it needs to create a truly customer-oriented organization by focusing on cost-efficient and customer-oriented product development as well as efficient coordination of world-wide demand with manufacturing and supply. For this reason, the case company has developed a process for customer-focused product development entitled Product Creation Process (PCP), which is detailed in the next section.

4.1 PRODUCT CREATION PROCESS

The PCP is a holistic process for managing product development. It describes all areas of creating products, while keeping the customer needs in focus. The purpose of the process is to make sure that all the necessary information is collected at the right time to realize efficient and effective planning in the short and in the long-term. The process is executed on product category level and it is project oriented, meaning that each product category is dealt with separately and that each time the process is initiated a new project is created. All the projects within a certain product category are managed by a business manager that is responsible for the progress of the projects. The PCP consists of four phases, which all contain steps, activities and check points that need to be dealt with during the product development work (Figure 1). The objective of the first phase (strategic market planning) is to ensure clear identification and prioritization of opportunity areas by establishing a strategic market plan for the product category. The plan is built on corporate prerequisites (e.g. product innovation strategy, brand and design strategy as well as global needs) along with industry analyses. Examples of questions that pertaining to this step are: on which areas should we focus our innovation work, which changes in consumer behaviour can create business opportunities, where are the growth markets, and what can we do that our competitors have not yet done?

The second phase consists of two parallel steps: Business opportunities and Primary development. During the business opportunity step an understanding of customer needs in the targeted product category is obtained through customer insight. To gather customer insights the case company uses several techniques such as observations, surveys and evaluations. However, observations are preferred, as observed behaviour is richer than described behaviour. Still, this kind of research is quite hard to employ since local product creation processes develops global products and the requirement can differ greatly between different regions or even between countries within a certain region. After that, the case company tries to identify means of developing products that satisfy the identified needs. Then it describes the needs and suggests a solution in a project proposal (PP). Each product category has a product council that decides whether proposed projects are accepted for product development or not. In primary development, technical solutions within the targeted product category are developed producing verified ideas or hardware solutions that can be applied to relevant projects in product development.

The objective of the third phase (product development) is to develop products based on an accepted project proposal. The phase consists of three steps, each including a number of stages (Figure 2). Each stage includes several
New Product Development in a Manufacturing Company – A Challenge for Supply Chain Management

activities that need to be dealt with, and before the project is approved to advance to the next stage it must pass a check point.

Figure 2: The Product Development Step

The objective of the first step (project specification) is to create a specification of requirements based on the project proposal and later conduct several theoretical investigations concerning how the specification may be realized. In this step as much information as possible is collected to provide an efficient front-loading of the process. This strategy is defensible since the projects in this phase still are rather inexpensive. Following this, product specifications are verified with respect to the collected information and the theoretical investigations. This means that one project proposal can result in more than one product development project. The last stage in the project specification step is prototyping, which refers to the creation of a prototype of the verified product specification. The prototypes are very precise, and several tests are conducted on them. These tests are conducted by cross-functional teams consisting of persons from several departments such as procurement, production, service, laboratory, R&D and product development. Two of the areas that are highlighted in these tests are ease of assembly and ergonomics. Exhaustion and firmness tests are not conducted in this stage since each prototype approximately cost 0.5 million SEK to develop.

If the prototype is accepted the project enters the next step, project industrialization. The length of this step differs between product categories. In the industrialization step, the case company needs to invest in real machines and tools. These investments can be quite large and it is thus important to run the industrialization step as efficiently as possible, regarding both cost and time, in order to provide revenue as fast as possible. The first stage of the industrialization step is to manufacture a couple of test batches of the product. Between each test batch the manufactured products are evaluated in several ways to identify errors and possible improvements. For instance, the products are tested in a laboratory (emissions, vibrations, noise) and in relevant field studies. A chainsaw for the North American, South American, and Scandinavian market needs to be tested in all these environments. After a number of test batches has been manufactured and no more errors or possible improvements can be identified, the product is approved for the next stage. In the next and final stage of the industrialization step, a number of pilot batches are manufactured. In this stage the purpose is to make sure that the company’s and its suppliers processes are aligned and working efficiently.

After the case company has secured the processes, the project is approved for the final step, production. This means that the company now starts to produce and sell the product on the markets. In the beginning of this step the case company continuously evaluates the results to determine if changes or improvements are needed. During this evaluation period the company also evaluates whether the project was successful or not. Finally the project is closed and production continues.

The objective of the final phase (phase out) is to ensure that the product assortment is updated accordingly to products life cycles and that obsolete products are properly out-phased. Both these objectives rely on a consistent follow-up period.
4.2 Roles and Responsibilities in Product Creation Process

Essentially three departments are involved in the PCP: Marketing, Product development and R&D. The product development department consists of several product divisions. Depending on where in the PCP a project currently is the involvement of the different departments differ (Figure 3).

Figure 3: Roles and responsibilities when executing Product Creation Process

The marketing department primarily works with the strategic market plan; product development primary works with business opportunities as well as generation and portfolio planning; while R&D primary works with primary and product development. The marketing department is closest to the customers; however, the product development department also work with customers regarding product concept and different investigations.

4.3 Results and Improvement Areas

After implementing the PCP the case company has improved the efficiency and effectiveness of product development. Improved effectiveness means that it to a larger extent develops product that their customers demand, while improved efficiency means that it develops more products faster and at a lower cost than before. The intensity of products development has increased, leading to more active product development projects. On a yearly basis the case company manages around 6-7 larger projects, each resulting in 5-6 new products, along with approximately 25 smaller projects concerning minor improvements, face-lifts and cost reductions of existing products. Moreover, the development time has been reduced over all product categories. Currently a professional product takes approximately 3.5 years to develop while simple consumer products take one year to develop.

The PCP focus on products attributes and the objective is to develop products that the customers desire. Questions concerning how to sell and introduce products on the market as well as how to distribute them to customers are not considered (i.e. customer service). These topics are often considered afterwards, by other personnel in the organization. The case company is aware of the need to investigate this issue further and considers changing the process in the nearby future. This is because product competitiveness is not solely restricted to product attributes, but also as it concerns customer service; these issues needs to be addressed in parallel. This is particularly the case in the consumer segment, where it is important to evaluate and provide accurate procurement and logistics alternatives (e.g. perhaps it is necessary to provide both ordinary shops and e-shops, requiring different logistics capabilities). This issue also becomes increasingly important since the consumer segment is increasing in comparison to the professional segment. One example where the case company was affected by the issue highlighted above was when it developed the Automower (an automatic lawnmower). When this consumer product was developed and ready for production, the case company realized that it was quite different to existing consumer products, and therefore required other distribution channels. These channels were not developed since this issue not had been encountered before. This led to an initial use of inappropriate distribution channels. The major issue was that this consumer product was more complex than ordinary consumer products and required skilled sales personnel, which the current consumer distribution channels did not provide.

5. Discussion and Conclusions

The conventional approach to product design and development is to assign representatives from support functions to review and recommend changes as the process evolves. The more recent concept of concurrent design involves a multi-function design team, which is highly structured and with greater responsibility and authority. Concurrent design has, so far, mostly focused on internal collaboration. However, in today’s global competition, concurrent
design, based on integration and collaboration in the extended enterprise is a vital key to success. It means that the design function is aligned and integrated with other main functions in the company and in the extended enterprise. Traditionally, most organizations have been functional in their structure with responsibility for each stage in the SC, including design, being separate from the other. Typically in these businesses the new product development process is linear and sequential with a consequent lengthiness of time-to-market and time-to-volume. In today’s challenging markets these ‘silo’ type structures are not capable of rapid response to fast-changing requirements. In the market places where short product lifecycles are the norm, delays in bringing products to the market can have detrimental consequences to sales and profit.

To sustain attractive in the future, businesses will need to produce innovative, high quality, highly value-added products and services and bring them quickly and effectively to the market. Two major issues need to be addressed: (1) the need to develop innovative, value-adding products; and (2) the necessity of bringing them quickly to the market. This can be achieved by simultaneously working with customer-oriented product development (to create an effective and efficient product development) and SCM (to create an effective and efficient SCs); which is a subject addressed in the emerging Demand Chain Management (DCM) approach. It is important to note that companies, in order to be successful, not only need to work with these areas simultaneously, but also connect them to one another. For instance, SCM is affected by product development; without new products, market acceptance and value added packages, an efficient SC is useless, supplying the wrong products very efficiently. SCM also affects product development since product competitiveness is not solely determined by product attributes, but also by the included services in the total value offering. Companies offer more than just a product, they offer a value package and therefore it is important that product innovation is not restricted to product attributes; it should also be applied in other areas such as customer service. This means that companies, when gathering information regarding needs of new products or product features, should also collect information regarding service needs and should consider integrating manufacturing and logistics in order to identify the most appropriate SC strategy for each customer or group of customers. Additionally, product development needs to be integrated with SCM in order to successfully introduce products on the market and to ensure that the product assortment is updated according to product life cycles and that obsolete products are properly out-phased.

This means that effective implementation of the product development process not only enables management to coordinate the efficient flow of new products, but also assists to support ramp-up of manufacturing, logistics, marketing and other related activities supporting the commercialization of the product. For this reason, companies need to address all these issues in parallel to be successful and this requires some type of integrative product development approach. Based on the most recent articles that has addressed design from a SC view [23-27] we propose the following definition for this integrative product development approach: “Design for the SC referring to design as a key element of the product development process and is concerned with designing the product, whilst taking into account the impact on the performance and success of the SC”. This approach requires a better understanding of the design process and the way it is integrated with other key functions in the company and in the extended enterprise. It is often stated that the SC starts on the drawing board. However, today’s global competition requires that the SC starts with impulses to new products and services coming from the customer insight process. For example, impulses for innovative product design at Electrolux come from the “Spark process”, which is part of their consumer insight process. Electrolux Product Flow Management is a kind of concurrent design approach and takes a process-oriented view of the whole flow “from the cradle to the grave”. This means that several functions are involved in a concurrent approach to design, development and innovation [28-29]. Hence, the entire process from customer insight to the phase out of the product has to be taken into account. The process has to be looked at from an end-to-end perspective. This, of course, cuts across a lot of internal and external boundaries. Boundary spanning was the focus of traditional business process management (BPM) approaches. However, the results have been rather meager. Success requires a shift of focus from simple mapping, simplification and coordination of processes to full scale cooperation and collaboration based on Information and Communication Technology (ICT). New tools, such as Service Oriented Architecture (SOA) and Event Driven Architecture (EDA) enable alignment and integration on operational, tactical as well as strategic levels [30]

This research shows how product development is structured and executed in an international manufacturing company, so that an innovative, predictable, and efficient product development may be realized. Case study findings reveal that the case company after implementing a strategic and structured PCP has improved the efficiency and effectiveness of product development. Improved effectiveness means that it to a larger extent develops products that their customers demand, while improved efficiency means that it develops more products, faster with a lower cost than before. The research has also shown that successful product development requires a holistic view from strategy to product design as well as profound knowledge of customers and their requirements. However, it can be concluded that the case company of this research work has not yet developed this kind of linkages between product development and SCM, since the PCP mainly focuses on products and product attributes. The major issue is that questions
concerning how to sell and introduce new products on the market as well as how to distribute them to customers (i.e., customer service) are not considered in the PCP. Still, the case company has become aware of this issue due to problems associated with the lack of integration between product development and SCM. Moreover, it has realized that this issue is increasingly important in the mature consumer markets in which it operates, where outstanding purchasing and logistics solutions is a success factor. An interesting aspect for further research is the above highlighted integrative product development approach. Two major issues should be addressed in such a research approach: (1) how do we define and develop innovative, value adding products? The answer to this question requires that we investigate how the customer insight process is performed and integrated in the design and development of products; and (2) how do we design, develop and manage responsive SCs with the design and development function as an integral starting point?

REFERENCES

New Product Development in a Manufacturing Company – A Challenge for Supply Chain Management


Quality Controls and Testing processes in Machine Tool Realisation

Francesco Aggogeri*, Marco Mazzola, Nicola Pellegrini
Department of Mechanical and Industrial Engineering
University of Brescia
Italy

ABSTRACT

The field of machine tool production is currently undergoing major changes due to increasing structural and dynamic complexity of the markets. In this context the quality control and testing processes have a fundamental role in order to provide a product that satisfies the customer need requirements. The main objective of a testing process is to deploy experimental verifications on products or components in order to establish the compliance with the design specifications. In this step it is important to follow specific procedures and activities in accordance with the main guide lines and international standards. The goal of this paper is to show and evaluate the main quality control and testing processes in order to realise a machine tool. Starting from the importance of quality concept, the authors analyse the different activities necessary to deploy an effective testing process. The paper focuses on the acceptance processes of raw material and components, the control processes of the main manufacturing systems (roughing, finishing and assembling activities), and final testing process of a machine tool. For every step the authors want to highlight the impact of the methods and procedures on the effectiveness and efficiency of a company underlining possible benefits and problems. The conclusions of this study explain a complete road map in order to implement quality control and testing processes for machine tool realisation.

1. INTRODUCTION

The main goal of this study is to describe a complete collection of testing processes that a modern organisation should adopt in order to design, realise and test a machine tool. The testing processes consist in a series of activities for checking the compliance of items/components with the design specification. In this way it is important to create an integrated flow (design, production and assembling processes) that develops quality control and testing activities efficiently [1]. These activities aim to guarantee that all products and components, that concur to create the final machine, are compliant with designer and customer requirements [2]. A “correct product” is a consequence of a series of managerial and practical activities. In particular, it is necessary to consider every specific product requirement from the design phase, mirror the functionalities of the product into the design and manage correctly production processes and quality controls [3].

The testing and quality control processes should be integrated in an effective flow, as shown in Figure 1. An integrated flow means that every function involved in a common process has to share a common language in order to achieve customer satisfaction. Through the process integration a shared way of thinking is spread and the importance of a goal widely understood in the company is emphasised. A fully integrated effective flow needs to consider the design of the product, the design of the manufacturing process and the design of the control process.

The main input of an integrated flow is the “expected product” (i.e. machine tool) that should be translated by design process in technical requirements. The manufacturing system follows the design guidelines for realising a final product that satisfies the customer needs. For this reason it is important to identify a package of tools that are able to guarantee and control the compliance between the requirements and the effective results obtained by the design and manufacturing processes. The design phase develops a deep study on the feasibility of the “required product” defining “if” and “how” the customer requirements can be satisfied in accordance with the manufacturing process owners. Furthermore a robust design should analyse all possible problems and failure effects identifying the main critical activities of the realisation process in order to define the main quality control tests. In fact, for a company perspective it is vital to comprehend causes and effects that could affect finished product and translate the analysis into the concrete design. The design and the related technical reports must offer a structure that considers every potential problem and difficulty that could emerge during realisation processes, from raw material acceptance to assembly and testing operations. The need of compliance between the “expected” product, technical design and

* Corresponding author: Tel.: (0039) 030-3715579; Fax: (0039) 030-3702448; E-mail: francesco.aggogeri@ing.unibs.it
“final” product can theoretically be reached adopting a Quality Management policy based on ISO 9001 standards that provide a series of requirements an organisation needs to fulfil in achieving customer satisfaction through consistent products and services.

Figure 1: The integrated flow in Machine Tool realisation

Following ISO 9001 [4] standards an organisation should define the quality policy understood and followed at every level and by every employee. Nevertheless, it is not enough to identify the main organisation goals and the improvement actions for reaching excellent results, but it is fundamental to study effective procedures that can be adapted to the integrated flow and guarantee the delivery of a product as well as it is expected by customer. This study wants to show practical tools and examples in order to develop testing activities in accordance with the Quality Management System of a company that realises Machine Tools.

2. THE ROLES OF TESTING PROCEDURES IN AN INTEGRATED FLOW

In order to show practical examples, in this paper the Authors have considered a modern company that realises Machine Tools (horizontal milling-boring machines, vertical milling machines and turning-milling machines). This company had implemented a Quality Management System following ISO 9000 standards [4, 5] and it well represents a typical context where testing and quality control processes assume a fundamental importance in customer requirement satisfaction. Figure 2 shows the main flow that manages testing processes. The main component is the Quality Management System characterised by three elements: ISO 9001 Requirements, Quality Manual and Procedures. These elements should define the quality control activities in the integrated flow related to design, material acceptance, manufacturing, assembling and final testing processes. In particular these procedures should monitor the main macro-processes and deliver instructions to every sub-activity.

What are the main goals of these instructions? Design testing procedures want to avoid a lack of comprehension of design specifications. In fact, often tolerances lack in updating, in comprehension and in affinity with product requirements. This problem can be filled only through a correct design definition based on the quality finished product perspective, measured during the last testing phase. In particular, the design testing procedures take into account lack of tolerances, information redundancy, lack of technical symbols, uncorrected position of tolerances or uncorrected specifications for symbols. The main goal of the material acceptance testing activities is to guarantee the compliance of the raw materials and components that are in-coming. Following ISO 9004 [5, 6] an organisation should also develop a procedure to define the quality level of suppliers based on their production and technological capability, certifications and reliability. The in-process testing considers all necessary control to verify the technical compliance of components that are realised inside the company, taking into account internal manufacturing and transformation processes. In the light of these results, it is important to define the main controls on assembling processes focusing on functionality of sub-groups and groups of machine tool.
The final step is the testing of Machine Tool (MT). This activity includes both the verification of the MT movements (translation and rotation axis) and the control of the MT output (i.e. part dimensions). For every phase it is fundamental to define effective and clear procedures that can be developed efficiently into the integrated flow. This paper wants to show some of these procedures using practical tools and examples.

### 3. Design Testing

In order to understand the design of testing activities it is necessary to introduce the concept of technical drawing. Technical drawing is the main tool for sharing information and data with production and testing departments. There are three main types of technical drawings: project drawing that describes the component functionalities imposed by the customer, production drawing that focuses on manufacturing processes and assembly activities, and testing drawing that verifies the functionalities of components.

Every drawing typology provides technical information, such as dimensional and geometrical tolerances [7]. The sharing of all data between the main processes, which identify problems and possible misunderstandings, is essential to increase the business performance. For example, the Authors consider a boring machine tool, as shown in Figure 3a. Through a functional decomposition, the MT can be divided into main groups: the boring machine head, the workbench, the ram, the head cart and the ram cart. In defining design testing, the first step is to identify functionality requirements of the machine tool head. Figure 3b shows these main functionality requirements: head movement on the cart, spindle line translation and rotation, adapter rotation and nut group translation and rotation, enabling the head translation. The nut group is external to the head and in particular the guide profile, hydrostatic plates and the nut support are essential for the working of the machine, as they impact on the movement accuracy. For example, the most important elements of the guide profile are the features of parallelism, planarity and orthogonality, because the guides are connected to the cart. This aspect can have a significant effect on the machine output, due to the influence of static and thermal deformations and vibrations on final results. The purpose of the nut is to get a parallel translation of the head. This aspect imposes features of parallelism and perpendicularity of the component.

The movement of the spindle line defines the MT accuracy. Therefore the testing requirements are:
- rectilinear movement on the w axis;
- perpendicular movement of the spindle on the w axis compared with the cart movement on the y axis;
- perpendicular movement of the spindle on the w axis compared with the cart movement on the x axis;
- roundness of the spindle rotation;
- periodic axial movement.

The definition of tolerances and testing of the spindle line group is complex. A shaft gives the motion transmission from the motor to the spindle line ($\Phi=160\ mm$; $L=2070\ mm$); flexions and distortions can imply an uncorrected functionality of the MT. Requirements for the shaft are parallelism tolerances (in particular to guarantee a correct connection between the cogwheels) and alignment. In order to satisfy these requirements it is possible to
use particular bearings. Their position is unreachable due to component geometrical features and is not verifiable in the testing process.

Figure 3: Overview of functionality requirements of the machine: milling boring MT (a) and MT head (b)

The design testing activities should be developed in order to verify technical tolerances, clarity of information, technical symbols, corrected position of tolerances or corrected specifications looking at group functionality. These aspects seem to be simple to understand and solve, but they are often neglected, as highlighted in Figure 4. In this case, the technical drawings lacked of dimensional tolerances, because the designer did not consider them as fundamental. Nevertheless it can assume a relevant importance in manufacturing or assembling processes. This is a typical consequence of incomprehension between departments. Using an effective testing procedure it was possible to avoid it and prevent a drastic problem in the integrated flow.

Figure 4: Missing dimensional tolerance (a) and relative solution (b) due to an effective Design Testing

4. ACCEPTANCE PROCESSES OF RAW MATERIAL AND COMPONENTS

The acceptance process aims to verify the compliance between the features of in-coming raw materials/components and designer requirements. A collection of procedures should check all critical parts (i.e. batch size, material features, dimensions, documentations) in order to avoid a reduction of efficiency or effectiveness in the integrated MT building flow. For this reason, it is necessary to plan different activities that include both simple inspections (i.e. number of pieces, type of documentation) and more complex controls (i.e. measurement of geometrical tolerances). These procedures contain a list of operative information related to type of supplier, non-compliance reports, testing observations, supply contracts, involved manufacturing processes and supply-chain.

In order to develop an effective testing plan it is possible to divide the acceptance process into sub-processes (namely Brick Process). Following this way, it is easier to identify the process owners and define operative activities. For example, the first step of an acceptance procedure is the verification of the compliance between in-coming materials and relative documentations (Brick Process 1 – BP1). Only when every order line is checked, it is possible to store in-coming materials. In case of nonconformities, an operator has to fill a report (NCR) and the “bad” components are stored in a dedicated area (Brick Process 2 – BP2), where the Quality Management Department analyses the rejection causes (Brick Process 3 – BP3). A significant part of in-coming items are checked using attribute and variable sampling plans in accordance with ISO 2859 (Brick Process 4 – BP4) [8]. By considering raw
materials it is important to check the compliance of documentations, material certification and basic dimensions (i.e. surface, rectilinearity). Instead, for every MT functional item the acceptance procedure requires a special test (*Brick Process 5 – BP6*), focusing on basic dimensions and geometrical tolerances using mechanical measurement systems (i.e. CMMs – Coordinate Measuring Systems). CMMs are accurate instruments used for precision inspection in industry, nevertheless there are still inaccuracies and uncertainties associated with them [9] due to a number of factors, such as e.g. measurement task, environment, operator and measurement procedures. In order to obtain reliable measurements, it is necessary to calibrate the machine following a number of activities [10, 11]. For example, these activities deal with the use of CMMs for comparative measurements with respect to a calibrated artefact equal in characteristics to the measured object. These operative steps should guarantee that all incoming items satisfy the designer requirements and effective assembling in the integrate flow.

5. In Process Testing and Quality Controls

5.1. Manufacturing Testing Activities

In-process testing and quality controls are related to internal company manufacturing (roughing, finishing) and assembling processes. It is possible to divide these controls in two main categories: self-controls and testing-assisted controls. In self-control activities, the operator should be able to decide directly (based on measurements) and/or indirectly (based on estimates) on the acceptance of an item. Self-control is performed on-machine using reliable equipments. Otherwise, the testing-assisted control is more complex and it developed in dedicated testing area. For every realised item there are specific procedures that require the verification of basic dimensions and geometrical tolerances. The main goal of these procedures is to guarantee a reduction of risks related to reworking and repositioning. In particular using brick process criteria, it is possible to find testing procedures related to preliminary controls (BP11), roughing processes (BP12), intermediate working processes (BP13), finishing (BP14) processes and repair (BP15).

Preliminary controls are applied to incoming raw pieces. A sample is tested to verify chemical composition, type of treatment, hardness and goodness. In case of conformity, the raw piece follows the roughing flow. It is set on MTs and positioned in accordance with the designer’s recommendations and the surface markers. At the end of the roughing processes, the intermediate and finishing processes transform a raw piece into a product. These steps are fundamental; in fact a working mistake could dramatically compromise the MT building activities. For this reason a number of measurement (basic dimensions and geometric tolerances) tests are performed on the board (BP16) of MT or in the metrological laboratory (BP17). These quality controls are also fundamental to define the performance level of the company manufacturing system, suggesting improvement actions and providing an important feedback. In order to develop a standardised method shared between every operator, it is necessary to make use of logic schemes adapted to every single component. In this way the testing operations can be made as linear as possible. This is achievable only using clear and appropriate testing reports and procedures. The testing forms have to be based on specified standards, thus enabling a comparison between expected or admitted values of the considered dimensions and the verified ones, as shown in Figure 5.

![Figure 5: Dimensional control format (a) and guidelines for dimensional control (b)](image)

The testing procedures for the evaluation of the quality standard compliance must include three important concepts: correct position of the analysed item/component, location and control of the references reported on the technical drawings, and control of all geometrical elements defined during the design.
The manufacturing processes also include the reparation and reworking activities. In fact it is possible to rework those products or components that a quality test has defined as rejected items (BP19).

5.2. **SUBGROUP AND GROUP ASSEMBLY PROCESS**

The assembling process of subgroups is defined by three sequential activities: the picking of components from the storage (B20), the assembly process (B21) and the checking of subgroup/group functionality (B22). The documentation includes the technical drawings of components and groups, the necessary equipments and the procedures related to the assembly cycle. The controls are generally performed by the operators.

The effectiveness of assembly process can be influenced by a number of factors such as group functionality (mechanical, electric, hydraulic, pneumatic), component tolerances, process difficulty and operator experience. For this reason it is necessary an effective definition of a collection of quality tests. For example, as shown in Figure 6, a quality test highlighted the impossibility to assemble component A with component B, even if the two items were complying. In fact the centre of M12 hole was out of position related to the centre of $\emptyset 22$ H7. Only an operative process of assembling could show this problem. For this reason it was necessary a review of technical drawings to add new tolerance reference (component B) and modify the testing control before the assembly process.

![Figure 6: Assembly process of a subgroup](image)

This example wants to show the importance of knowledge of component functionality. The integration between company functions and the sharing of information, data and procedures is the starting point for “correct” products realisation. The last step of a MT building flow consists in assembling groups. This process follows the same guidelines adopted for the subgroups even if there is an increase of MT component sizes and complexity. This process is fundamental for the functionality and accuracy of a MT. For this reason a strong interaction should be emphasised between assembly and testing processes.

6. **FINAL TESTING PROCESS OF A MACHINE TOOL**

The final testing process consists in a single procedure that includes the preparation of the MT, the testing of geometries and operations, and further phases of adjustment. The final testing can be defined as well as the phase of MT “tweak” in accordance with customer requirements. This process should be performed by an expert and heterogeneous team. The testing procedure must be able to verify the MT performance, make operational and service changes, and set the MT to repair any further non-conformity. All controls must follow rigorous guidelines to ensure and demonstrate the validity of the measurements. The process can be divided in five steps: adaptation to environmental condition (BP23), positioning of the machine tool (BP24), testing of geometries (BP25), testing of manufacturing features (BP26) and further adjustment and compensations (BP27). In order to satisfy the first procedure, the MT should be tested in appropriate environmental conditions. For example, the temperature should be maintained between 18°C and 22°C. A variation of $\pm 2$°C is permitted for the 12 hours prior to the test, in order to eliminate the thermal inertia of the parts. A correct position of MT avoids problems related to the verification of the main axes (parallelism or perpendicularity). The geometric testing process is the longest and most laborious step. Often it is necessary to repeat the tests many times in order to satisfy the required compliance. In this phase it is fundamental to demonstrate the accuracy and repeatability of positioning of the axes of the machine (UNI ISO 230-2, [12]). Figure 7a shows a useful format for checking the performance of a linear axis, referred to the general axes scheme shown in Figure 7b. In this case all features are checked considering a comparison between the verified
values, nominal values, admitted values, defining a coordinate system.

![Figure 7: Guidelines for linear axis testing (a) and scheme of axes testing (b)](image)

If the MT has been successfully tested in geometry, the final verification phases consist in the controls of manufacturing performance levels (i.e. milling, boring, turning) and the potential integration with other working station flows (in case of FMS modules).

The assembled parts should be measured in a metrological room in accordance with UNI ISO 10791-7 [13]. The process of adjustment and repair (BP26) is associated to every single test operation, because it is often rare every group of the MT can immediately satisfy an accuracy of microns. The process of adjustment and repair involves different figures, such as the tester, the assembly manager, the quality manager and the designer. Only a complete analysis of documentations, technical drawings and quality tests can suggest possible solutions and improvement actions.

7. CONCLUSIONS

In the light of these considerations it is possible to conclude that the quality controls and testing processes should be integrated in the main MT building activities (i.e. design, material acceptance, manufacturing and testing). This paper wants to demonstrate the importance of effective procedures in an integrated flow. Their goal is not only to verify the compliance of quality standard levels but also to provide an important feedback in order to improve the operational processes. In fact, if every decision about design features comes from an integrated planning team, involving both design, manufacturing and testing functions, it is assured that the finished product successfully reflects every customer requirement, both internal and external. To provide a machine tool that satisfies the customer quality standards, a good knowledge of quality ideas and a widespread sharing of main management tools are absolutely needed. In this context, the first step is a review of technical drawings that are the mirror of problems and misunderstandings in manufacturing process activities. Technical drawing appears as a fundamental tool to integrate information and data with production and testing processes. Based on real examples, the paper highlight different quality control and testing procedures that should be performed to guarantee an effective and efficient MT building flow. A component of a product is realised through different phases (idealisation, design, manufacturing and testing). To manage these activities in a successful way, an organisation should considerate all steps in an integrated manner rather than a sequence of operations, looking at the main product objectives. If the cooperation and the agreement of company functions lack in the understanding and in the definition of the product goals, it could be difficult, for example, to execute a work or to define tolerances, or to perform final controls when a technical drawing is defined for design purposes only.
Using a brick process criteria it is possible to discover testing process owner and define a common strategy to fit all organisational activities. A Quality Managements System can suggest a generic guideline to ensure a product that satisfies customer requirements, nevertheless only the analysis of problems permits an effective implementation of improvement actions. For every step a testing procedure should be performed, by sharing a common language with the other departments and processes. This language is represented by technical tools and quality reports. In particular the quality reports must reflect every problem related to a component/item. This concept is a milestone of an integrated approach.

ACKNOWLEDGEMENTS

This paper is dedicated in fond memory of Prof. Enzo Gentili.

REFERENCES

Using Statistical Process Control to Monitor Inventory Accuracy

John R. English, Ph.D., P.E., Kyle Huschka, and John M. Miller

ABSTRACT

Inventory accuracy is critical for almost all industrial environments such as distribution, warehousing, and retail. It is quite common for companies with exceptional inventory accuracy to use a technique called cycle counting. For many organizations, the time and resources to complete cycle counting are limited or not available. In this work, we promote statistical process control (SPC) to monitor inventory accuracy. Specifically, we model the complex underlying environments with mixture distributions to demonstrate sampling from a mixed but stationary process. For our particular application, we concern ourselves with data that result from inventory adjustments at the stock keeping unit (SKU) level when a given SKU is found to be inaccurate. We provide estimates of both the Type I and Type II errors when a classic C chart is used. In these estimations, we use both analytical as well as simulation results, and the findings demonstrate the environments that might be conducive for SPC approach.

1. INTRODUCTION

Inventory accuracy is critical for most industrial environments. Most large and successful retail and warehousing environments desire to maintain accurate perpetual inventory. To achieve accurate inventory, the recorded inventory for stock keeping units (SKUs) must match the actual on-hand inventory. If inventory accuracy is low, there are many problems that can arise. First, there are non-value added activities that must take place to rectify the discrepancies of the inventory. These tasks include correcting an inventory data base, hurried replacement of product, and any other time taken by employees to correct the inaccurate inventory. Although these costs are reason enough to validate the importance of accurate inventory, the most significant impact of inaccurate inventory is the loss of customer good will, and it occurs when there is insufficient inventory to satisfy customer demand. Alternatively, time can also be wasted on unnecessary replenishment of inventory.

Cycle counting is a proven method to monitor and check inventory accuracy on a continual basis. Today, cycle counting is often used as a replacement of annual physical inventory checks, which require the entire factory or warehouse to be shut down to count all SKUs at one time. Basically, cycle counting is a process in which SKUs that are kept in inventory are counted periodically to examine if on-hand inventories match the inventory recorded in the data base. If discrepancies are found, the reason for the discrepancy is examined, and the inventory records are updated, or inventory adjustments are made. Most methods of cycle counting require counting all SKUs over a period of time to ensure that the actual inventory matches the inventory that is recorded in the entire company’s inventory data base. [1] states that “through the proper use of cycle counting, inventory record accuracy above 95% can be consistently maintained”.

However, the feasibility of cycle counting can become controversial when considering large retail or warehousing environments that contain thousands of SKUs. With such large environments, the labor requirements to implement cycle counting could become excessive. For this reason, large retail and warehouse environments pose a problem for the implementation of cycle counting.

It has been determined that inventory accuracy is important to companies as it creates a significant value to the operations. However, achieving accurate inventory can cost a large amount of money and labor. A new, cost effective method for monitoring inventory accuracy for large retail and warehouse environments needs to be developed.

Statistical Process Control (SPC) is an effective statistical approach used to monitor processes and improve quality through variance reduction. Basically, SPC uses sampling to monitor the performance of a process. An advantage of SPC is that it requires reduced resource expenses, because it depends on random samples and not 100 percent inspection. It is obvious that the aforementioned problem of finding a cost effective method for monitoring inventory accuracy for large operating units would be a suitable application of sampling approaches. If correctly designed, SPC could be used to monitor the inventory accuracy of a warehouse without 100 percent inspection. There are two specific tools that would be suitable for this environment.
First would be the use of a p chart, which is used to monitor the percent nonconforming of inventory accuracy for a given operational unit, as demonstrated in [2]. For the p chart, a random sample of n SKUs is selected, and each SKU is examined to determine if it meets the quality requirements. The observed number of SKUs where physical inventory doesn’t match recorded inventory is divided by the sample size to provide an estimate of the inventory accuracy, or p. The observed p’s are plotted in relationship to time, and their variability is assessed. If non-random behavior is observed, process adjustments are pursued to rectify the problem.

The second option would be the use of a c chart, which is used to monitor the number nonconformities in a process. For a c chart, an inspection unit of size n is sampled, and the observed number of nonconformities is plotted in relationship to time. Similar to the p chart, the observed c’s are analyzed for non random behavior. A c chart could be used for monitoring inventory accuracy to examine the number of inventory adjustments for a given operational unit. The use of a c chart is the focus of the research in this paper.

Historically, SPC and cycle counting are domains that have not been considered to be complementary of one another. If these two methodologies can be incorporated, a cost effective method to accurately monitor inventory for large operating units can be realized. This paper shows that SPC can be applied to environments needing to monitor and/or improve inventory accuracy, specifically, the number of inventory adjustments for a given operational unit. It is the thesis of this work to present the c chart as an efficient alternative to cycle counting for monitoring inventory accuracy.

This research focuses on examining a population, which is based on a large retail or warehousing environment that consists of thousands of SKUs. The likelihood of a given SKU being accurate at a given time will follow a Bernoulli process. Each SKU within an operational unit could behave uniquely carrying their own Bernoulli rate; therefore, the basic assumptions of SPC are violated (e.g. each sampled unit follows the same distribution). It is the premise of this effort that such populations can be modeled by underlying sub populations that follow a mixture distribution. We develop and analyze a SPC approach for cycle counting by examining the use of the c-chart for populations that follow a mixture distribution. The experimental units measured in this c chart are the number of inventory adjustments made to a specific SKU. The number of adjustments made to the inventory records is being utilized as these data are typically monitored in most organizations. We will inspect the type I and type II error rates of the c chart with such populations. In the paper, we present the modeling fundamentals, and the results are to be presented at the conference.

2. BACKGROUND LITERATURE

The purpose of this research is to illustrate how SPC can support an improved method of monitoring inventory accuracy over cycle counting. In this section we introduce cycle counting and the various methods of cycle counting being employed today. We also give an overview of the SPC tools that are necessary for this research.

2.1 OVERVIEW OF CYCLE COUNTING

Cycle counting in essence is the process of counting inventories on a regular basis throughout the year. The use of cycle counting has become increasingly popular and has begun replacing annual physical inventories. Performing an annual inventory check requires shutting down the entire factory one or more times a year and conducting wall-to-wall physical inventory checks. Basically, cycle counting is a method that replaces annual inventory checks with a process that involves counting a small portion of inventory on an incremental basis with the objective of inspecting each SKU within an operational unit during a fixed period.

The process of examining portions of inventory with cycle counting involves sampling. Specifically, certain SKUs within a given inventory are selected, and the accuracy of those SKUs is measured. Once this has been done it can be inferred that the accuracy is a characteristic of the entire population. The ability to make this inference, opposed to surveying the entire population, allows cycle counting to save both time and money when it is implemented. The primary goals of cycle counting are to identify the causes of the errors, correct the conditions causing the errors, maintain a high level of inventory record accuracy, and provide a correct statement of assets [3].

[1] states that “with the proper use of cycle counting, inventory record accuracy above 95% can be consistently maintained.” [4-5] Portray cycle counting as a quality assurance process that emphasizes the correcting of errors.

There are various approaches of cycle counting including the geographical method [6], random sampling [1], the ABC method [7-9], and process control [1].
2.2 STATISTICAL PROCESS CONTROL

The use and understanding of SPC tools is inherent for this work. [10] states that “The field of statistical quality control can be broadly defined as those statistical and engineering methods that are used in measuring, monitoring, controlling, and improving quality.” SPC is basically the use of statistical techniques to measure and analyze the variation in processes, and improve the quality of the process through variance reduction. The four basic steps of SPC are measure the process, eliminate undesirable variation from the process, monitor the process, and continuously improve the process. There are seven common tools that are used in SPC: Check Sheet, Pareto Chart, Flow Chart, Cause and Effect Diagram, Histogram, Scatter Diagram, and Control Chart. The fundamental and most commonly used tool of SPC is the control chart, and it is the tool that is being examined in this paper.

The control chart was first proposed in 1926 by [11-12]. [13] states that all processes are subject to degrees of variation, which explains the basic principal of why the control chart is a useful monitoring technique for almost all processes including inventory accuracy. The chart has a center line (CL) and upper and lower control limits (UCL and LCL). The center line represents the mean of the process characteristic when the data collected are from a stationary process. The LCL and UCL indicate a range in which practically all of the data points need to be within in order for the quality characteristics to be considered in control. Control charts are widely used for both variable, as well as, attribute data.

2.3 TYPE I AND TYPE II ERRORS FOR CONTROL CHARTS

The control limits on control charts are typically drawn at 3 standard deviations away from the mean of the underlying population. Whenever a data point falls outside of these limits, it indicates an out of control condition. There are two types of errors associated with control charts: type I and type II. Type I or alpha errors are known as “false positives” and occur when we reject the null hypothesis (H₀) but it is actually true. Plainly speaking this is when a difference is observed but in truth there is none. In the application of control charts, a type I error is made when the operator concludes the process is out of control when in fact it is in control. Type II or beta errors are known as “false negatives” and occur when we fail to reject H₀ but the alternative (H₁) is actually true. In other words this is the failure to observe a disorder in a process. In the application of control charts, a type II error is made when the operator concludes that the process is in control when it is in fact out of control.

The average run length (ARL) is the average number of points that is needed to detect an out of control condition. It is simply the inverse of the probability of detecting a given process condition. Large in-control ARLs and small out of control ARLs are desired. ARL calculations can become quite complicated as documented in the literature. For example, [14] model control charts using a Markov chain approach to evaluate the ARL when supplementary run rules are applied. [15] present a model of the synthetic control chart and use it to evaluate the zero-state and steady-state ARLs. [16-17] look at using the EWMA control chart instead of the X̄ chart as well as, how to evaluate different methods of computing the ARL properties of EWMA control charts. [18] studies the effect of using run rules on X̄ charts and determines that they improve the sensitivity of the charts, but cause the chart performance to differ more from the known parameter. As mentioned earlier [14] use Markov chains to compute the ARLs for the X̄ chart, but their methods are general and can also be applied to other control charts. Finally, [19-20] extend the study of the X̄ chart with variable parameters to include variable action limits. Many additional research advances in the determination of ARLs for most all charts have been made and can be found in the literature.

3. MODEL DETAILS

In this work, we concern ourselves with the need to track inventory accuracy and it is a comprehensive extension to the concepts presented in [21]. In many retail and warehousing environments, actual inventory accuracy (e.g., the ratio of units on the shelf divided by the number of units thought to be on the shelf) at the SKU level may likely not be monitored. Alternatively, many companies will simply track a SKU as being accurate, that is, the number of units on the shelf match the number of units thought to be on the shelf, or inaccurate. As a result, a simple 0 or 1 will be documented for each SKU within a department and the resulting percentage of SKUs that are accurate for a given operating unit is tabulated. In [22], the authors model the use of a classic p chart to monitor such data. Guidance on sample sizes is presented and resulting Type I and II errors are presented.

In this work, we extend [22] to provide a basis to monitoring data that are commonly collected in lieu of tracking the percentage of SKUs that are accurate for a given operating unit. Specifically, the financial impact of good or bad inventory accuracy is objectively measured by the number of inventory adjustments that are made for a given operating unit. When a given SKU is found to be inaccurate, actions will be deployed to adjust current inventory
levels. That is, units may be added to the shelves when shortages are discovered, or the number of units recorded in the asset data base may be reduced when the shelf counts exceed the recorded on-hand inventory. Regardless of the adjustment, physical inventory adjustments are required. Such data must be monitored as the asset value of a given operation is adjusted.

The resulting data stream resulting from such an environment is a set of inventory adjustments recorded for each SKU. It is also necessary to understand that various SKUs within an operational unit are stochastically unique. It is fair to expect the number of inventory adjustments for a given SKU to follow a Poisson distribution with a rate \( \lambda_i \), but it is not feasible to assume that each SKU will follow the same distribution. It is our thesis, that at best, a cluster of SKUs will follow the same Poisson distribution. As a result, we propose that the random number of inventory adjustments, called \( X \), for a given large retail or warehousing unit can be represented by a finite set of subpopulations. That is, the probability density function (pdf) describing the number of inventory adjustments follow a mixture distribution that is shown in equation (1)

\[
\begin{align*}
 f_x &= p_1 f_{y_1} + p_2 f_{y_2} + \ldots + p_n f_{y_n} \\
 &= \sum_{i=1}^{n} p_i f_{y_i}
\end{align*}
\]

(1)

Where: \( p_i \) is the proportion of the population following the Poisson distribution with rate \( \lambda_i \), and \( f_i \) is the Poisson probability density function for subpopulation \( i \), and the proportion of the population follows equation (2)

\[
\sum_{i=1}^{n} p_i = 1
\]

(2)

The resulting expected value and variance of \( Y \) can be easily determined using basic expected value operators and is shown in [21].

Sampling from such a distribution is representative of what happens in our application domain where an organization desires to monitor the statistical stationary consistency of the inventory adjustments through random sampling. That is, at a given sampling period, a random sample of \( m \) SKUs would be the accumulated number of inventory adjustments required to rectify the inventory errors. The accumulated number of inventory adjustments from the sampling of \( m \) SKUs can be represented as random variable \( Z \) and is modeled as shown in equation (3)

\[
Z = X_1 + X_2 + \ldots + X_m
\]

(3)

Where: \( Y_i \) follows (1)

Clearly, the resulting distribution of \( Z \) is complex and challenging to derive for even the smallest values of \( m \). For our application, we suggest the plotting of such accumulated values of \( Z \) on a classic c chart where the mean and variances are based upon the mixture distribution of (1) stated above. In accordance to the findings of [21], the sample sizes can be upwards of 200-400 for operating units of reasonable sizes containing thousands of SKUs; therefore, the determination of various ARLs may be based on basic probability calculations for smaller values of \( m \), but for reasonable sizes of \( m \), simulation must be utilized. For the purposes of this paper, we present nested conditional probabilities for \( m \) of smaller values (2 and 3) that can be used for analytical determinations of the ARLs. At the conference, we will present some of the analytical findings, and we will extend the findings to practically suitable sampling schemes with simulation.

3.1 ANALYTICAL MODELS

When 2 random samples of SKUs are collected from the population described in (1), the total number of inventory adjustments for the sample is simply the sum of \( X_1 \) and \( X_2 \) (on \( n=2 \), using our notation). The resulting control limits are shown in equation (4)

\[
\text{Control Limits} = 2E[X_i] \pm 3\sqrt{2Var[X_i]}
\]

(4)

Using the UCL for a sample of 2 can be seen in equation (5),

\[
P[X_1 + X_2 \leq UCL] = \sum_{X_1=0}^{\text{UCL}} P[X_1] \times P[X_2 \leq (UCL - X_1)]
\]

(5)
Using the LCL for a sample of 2 can be seen in equation (6),
\[
P[X_1 + X_2 < LCL] = \sum_{X_1=0}^{UCL} P[X_1] \times P[X_2 < (LCL - X_1)]
\]
(6)

The probability of a type I error when \(n=2\) is shown in equation (7)
\[
1 - \alpha = P[X_1 + X_2 \leq UCL] - P[X_1 + X_2 < LCL]
\]
\[
\alpha = 1 - (P[X_1 + X_2 \leq UCL] - P[X_1 + X_2 < LCL])
\]
(7)

For \(n=3\), the control limits can be seen in equation (8)
\[
\text{Control Limits} = 3\mu[X_1] + 3\sqrt{3\text{Var}[X_1]}
\]
(8)

Similarly, the resulting \(\alpha\) from the UCL is shown in equation (9)
\[
P[X_1 + X_2 + X_3 \leq UCL] = \sum_{X_1=0}^{UCL} P[X_1] \sum_{X_2=0}^{UCL-X_1} P[X_2] P[X_3 \leq (UCL - X_1 - X_2)]
\]
(9)

And the LCL can be seen in equation (10),
\[
P[X_1 + X_2 + X_3 < LCL] = \sum_{X_1=0}^{LCL} P[X_1] \sum_{X_2=0}^{LCL-X_1} P[X_2] P[X_3 < (LCL - X_1 - X_2)]
\]
(10)

Therefore, using the probability of being within the range of the control limits, the type I error rate, \(\alpha\) is defined in equation (11)
\[
1 - \alpha = P[X_1 + X_2 + X_3 \leq UCL] - P[X_1 + X_2 + X_3 < LCL]
\]
\[
\alpha = 1 - (P[X_1 + X_2 + X_3 \leq UCL] - P[X_1 + X_2 + X_3 < LCL])
\]
(11)

4. CONCLUSIONS

In this paper, we present an alternative means of monitoring inventory accuracy in lieu of classic cycle counting. Although cycle counting is known to be exceedingly effective for improving inventory accuracy, many organizations often cannot afford to expend the resources required to conduct the necessary inventory assessments. We offer statistical process control as an alternative means to cycle counting. We demonstrate the utility of the c chart to monitor the number of inventory adjustments recorded for a given operational unit. When these findings are coupled with our previous work in [21], two reasonable alternatives to classic cycle counting are realized and their effectiveness in detecting significant changes is shown to be acceptable.

In application, large warehousing and retail environments are encouraged to utilize the SPC approach for monitoring inventory accuracy as the benefit of statistical sampling drives significant reductions in costs as both personnel are better utilized and inventory accuracy can be improved.
REFERENCES


Structuring stakeholders’ requirements concerning intra-logistic facilities in the context of QFD


Chair of Quality
Dortmund University of Technology
Dortmund, 44227, Germany

ABSTRACT

Numerous stakeholders and requirements are involved in the planning and developing of intra-logistical facilities and must be considered and implemented in solutions which will satisfy customer needs. To allow adequate requirements management these must first be captured and structured. A model has been created as part of the sub-project A1 of the German Collaborative Research Centre 696 which allows the requirements made of an intra-logistical facility to be structured. This paper will present excerpts of this model based on existing analyses and identification of stakeholders in an intra-logistical facility.

1. INTRODUCTION

In today’s global competitive environment, the significance of customer orientation, combined with cost pressure and the demand for higher rates of flexibility and shorter delivery and lead times, is steadily increasing. Companies are therefore searching for opportunities of saving in order to meet these demands. Particularly the field of logistics and intra-logistics open up potentials for optimization. Intra-logistical facilities are understood to be in-company material flow systems which serve to convey goods from the incoming goods area through to production and then to the goods dispatch area. [8] They include systems for storage management, material transport and order picking operations. Intra-logistical facilities place high demands on developers and manufacturers and incur high costs regarding investments and operation. However, use, maintenance, disposal and recycling processes and the thus resulting requirements of the facility are afforded only inadequate consideration in planning and purchase decisions. With regard to customer-oriented product design though, it is essential that these requirements are integrated as they can also contribute considerably to cost savings. Consideration must also be given to numerous areas of facility operation as well as to the facility manufacturer and the requirements of external parties, for example legal requirements, and implemented in the best possible selection or development of the product. The diversity of stakeholders and their numerous requirements which need to be taken into account concerning the facility which is to be developed and set up however make the implementation process difficult.

In order to find the best possible design for an intra-logistical facility for a special-purpose application, it is necessary to determine and integrate the requirements of all the stakeholders involved. Basically, the established QM method of quality function deployment (QFD) can be used to implement these requirements into respective solutions. The prerequisites for using this method are knowledge and weighting of customer requirements. However, when planning intra-logistical facilities, these requirements need to be structured in the run-up to a QFD in order to be able to handle the multitude of information. With the help of structuring models, requirements can be captured into groups thus allowing focus on specific requirements [2, 3, 4].

Part of the German Collaborative Research Centre 696 “Logistics on demand”[12] project A1 included the “Model for structuring and clustering requirements for intra-logistical facilities”. This looks at the development of a systematic structuring model for planning intra-logistical facilities to provide more customer-oriented products using a roll conveyor as an example.

* Corresponding author: Tel.: (0044) 1642-342482; Fax: (0044) 1642-342482; E-mail: FAIm2009@Tees.ac.uk
2. Necessity of Structuring Requirements of Intra-Logistical Facilities in the Run-up to Quality Function Deployment

Requirements made by various stakeholders, e.g. the customer, and the weighting of these requirements form the starting point of any QFD. When carrying out the QFD, these requirements are transferred into possible solutions and/or product features using a correlation matrix, in this case for intra-logistical facilities. Furthermore, conflict relations between single product features are investigated, a competitive comparison from a technical point of view and from the customer’s point of view is carried out and the technical implementation of the product features is assessed. In this way, the product and the characteristics of the facility are given the best (possible) design which will meet the needs of the customer [1].

The prerequisite for a successful QFD and translation of requirements into product features is therefore the existence of adequate information on existing requirements and their weighting. To achieve this, it is necessary to systematically capture and handle the required information; however, because of the diversity of stakeholders and the multitude of their requirements, the planning and development of an intra-logistical facility is difficult. To give a few examples: the requirements of maintenance personnel which include easy access to the facility and to the part which is to be maintained; requirements of the facility’s operator which might include low energy consumption and low acquisition costs plus as few costs and little time as possible to train facility personnel, and legal requirements regarding compliance with standards and regulations.

In addition, consideration must be given to the fact that requirements (may) display very different degrees of abstraction. For example, the requirements made of a roll conveyor might be that “it must be possible to expand the facility’s flexibility” or “it must be possible to service the facility whilst it is in operation”. Using the requirements during a QFD however assumes requirements are of the same, or similar, degree of abstraction otherwise the method will not produce any sensible results. It is therefore necessary to structure the requirements so that all those involved in the planning have a comprehensive overview of all existing information regarding the requirements and that not only gaps in the information are clearly shown but also focus can be placed on individual requirements and requirement categories. A suitable structuring model must be developed for this purpose. In the first step, stakeholders whose requirements are to be taken into account must be identified.

3. Stakeholders in Intra-Logistical Facilities and Their Requirements

Stakeholders are groups upon which a company depends for its survival, which are impacted by corporate activities and which are related to these activities [6, 7]. The following groups or persons are understood to be those who make requirements of intra-logistical facilities regarding the planning of such facilities. On the one hand, a difference is made between a manufacturer and an operator of such a facility and on the other hand between other external stakeholders such as the legislator, auditors, residents in the neighbourhood of the facility etc. Within the first two parties – manufacturer and operator – a variety of individual stakeholders can be identified; for example, management, purchasers, maintenance staff, developers, controllers, sales and marketing, operators. All of these stakeholders make many different requirements of an intra-logistical facility. For example, for management and controlling the focus lies primarily on financial issues such as low-cost acquisition and cost-efficient operations; whilst factors such as an ergonomically designed working place and easy operation of the facility are of more importance to the real “user” of the facility. In contrast, external stakeholders require compliance with laws and regulations, low levels of emissions and noise pollution, for example.

Considering the life cycle of an intra-logistical facility, it is clear that stakeholders and their requirements have an impact on different phases of the cycle. The manufacturer’s requirements are afforded priority upstream of those of the operator. When planning the facility however, all requirements must be taken into consideration independent of which phase of the life cycle they impact in order to be able to design a facility which will meet all of the stakeholders’ requirements. Taking requirements into consideration which have been ignored in the planning stage is extremely difficult and incurs high costs - the retrofitting of a facility to comply with emission and noise regulations, to name just one example.

The following diagram shows a comprehensive overview of stakeholders in an intra-logistical facility whose requirements must be determined and considered – the list is not exhaustive but reflects the current state of research.
In order to be able to adequately consider all identified stakeholders and their requirements in the planning of an intra-logistical facility it is necessary to have a suitable structuring strategy which allows all requirements to be captured and displayed. The following looks at and investigates existing structuring methods.

3. OVERVIEW OF EXISTING STRUCTURING METHODS

Literature describes many strategies on the topic of structuring [e.g. 5, 9, 10, 11]. A difference must be made between strategies which are not explicitly concerned with structuring but whose application requires structuring in the run-up. These include e.g. the FMEA or also QFD. In contrast, other strategies such as cluster analysis and the KJ method do not work with a predefined structure but by application of the method. The third group of methods structures use criteria or categories whose numbers differ greatly in the individual methods (cf. Figure. 2). Such strategies however cannot be transferred to the areas of application of intra-logistical facilities. Part aspects and individual structuring methods such as the product model in the St. Galler strategy may be taken over for the model which is to be developed but are not adequate for capturing and structuring the requirements in the area of application because of the complexity of the object under observation. Combination, adaptation and expansion of existing strategies are therefore required. This is the object of research in the sub-project A1 of the German Collaborative Research Centre 696. The current findings of this research work will now be presented.
<table>
<thead>
<tr>
<th>Structuring method</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility life cycle</td>
<td>Planning/ Sales</td>
</tr>
<tr>
<td>Work process intra-logistics</td>
<td>Implementation Use</td>
</tr>
<tr>
<td>DIN 2330</td>
<td>Consistency Relation Function</td>
</tr>
<tr>
<td>Ehrlenspiel</td>
<td>Technological and functional requirements Interfaces Costs Laws/standards/patents/guarantees</td>
</tr>
<tr>
<td>Time</td>
<td>Personnel Auxiliary means</td>
</tr>
<tr>
<td>Finchau</td>
<td>Utilitary Facilitarilly Ecological Economical</td>
</tr>
<tr>
<td>Hubke</td>
<td>Function Function-related attributes Operational attributes Ergonomic attributes Appearance attributes (aesthetic attributes,) Distribution attributes</td>
</tr>
<tr>
<td>Delivery and planning attributes</td>
<td>Legal compliance attributes Manufacturing attributes Construction attributes Cost-efficiency attributes Manufacturing attributes</td>
</tr>
<tr>
<td>Industrial design</td>
<td>Functionality Manufacturing Economy Ecology</td>
</tr>
<tr>
<td>DIN 31501</td>
<td>Servicing Inspection Maintenance Improvements</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Control Employees Deployment of system Warranty and guarantee Documentation Further development</td>
</tr>
<tr>
<td>Intervals/times</td>
<td>Operating systems Information systems Laws Equipment Monitoring</td>
</tr>
<tr>
<td>KANO</td>
<td>Basis Performance Enthusiasm</td>
</tr>
<tr>
<td>Kickermann</td>
<td>Component-specific Assembly-specific Product-specific Interface-specific</td>
</tr>
<tr>
<td>Company-specific</td>
<td>Sector-specific Universal</td>
</tr>
<tr>
<td>Krusche</td>
<td>Purpose Product feature Function variable Neighbouring systems Life cycle Company</td>
</tr>
<tr>
<td>Mannel</td>
<td>Administration Information management Finance management Energy management Material management</td>
</tr>
<tr>
<td>Personnel management</td>
<td>Production Sales R&amp;D</td>
</tr>
<tr>
<td>Myers Shocker</td>
<td>Characteristics Imagery Benefits</td>
</tr>
<tr>
<td>Pahl Beitz</td>
<td>Geometry Kinematics Forces Energy Substance Signals</td>
</tr>
<tr>
<td>Safety</td>
<td>Energy Manufacturing Control Assembly Transport</td>
</tr>
<tr>
<td>Use</td>
<td>Maintenance Recycling Costs Deadlines</td>
</tr>
<tr>
<td>Quality control loop</td>
<td>Market research Concept Draft Testing Manufacturing planning Procurement</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Final inspection Storage Dispatch Maintenance Disposal</td>
</tr>
<tr>
<td>Roth</td>
<td>Fixed requirement Target requirement Desired requirement</td>
</tr>
<tr>
<td>Sakowski</td>
<td>Spatial-operational Production engineering Technical IT Operational Personnel</td>
</tr>
<tr>
<td>Servqual</td>
<td>Reliability Assurance Tangibles Empathy Responsiveness</td>
</tr>
<tr>
<td>St. Gallen</td>
<td>Product core Extended product Formal product</td>
</tr>
<tr>
<td>Tanaka</td>
<td>Hard function Soft function</td>
</tr>
<tr>
<td>Ten Hompel</td>
<td>System-specific criteria Goods-specific criteria Organisational criteria Overall criteria</td>
</tr>
<tr>
<td>VDI 225</td>
<td>Number-relevant attributes Geometrical and kinematic attributes Mechanical attributes Thermal attributes Electrical and magnetic attributes</td>
</tr>
<tr>
<td>Optical attributes</td>
<td>Acoustic attributes Material and chemical attributes Production and assembly attributes Use, operation and maintenance attributes</td>
</tr>
<tr>
<td>VDI 2247</td>
<td>Wzh Requirement</td>
</tr>
<tr>
<td>Wögebauer</td>
<td>Operational tasks Realization tasks</td>
</tr>
</tbody>
</table>

Figure 2: Overview of existing structuring methods
4. MODEL FOR STRUCTURING REQUIREMENTS OF AN INTRA-LOGISTICAL FACILITY

A multi-dimensional model for capturing and structuring requirements is needed because of the large number of stakeholders and their requirements and associated complexity. By selecting the number of structuring axes and design of the respective sub-division of axes in categories areas can be created in which to order the requirements whereby the axis and the respective categories must be chosen in such a way to allow the respective requirements to be allocated to one intra-logistical facility. Care must be taken to avoid duplication and independency of the classes. Classes with different axes should therefore not be strongly similar to one another as when these classes are compared, no statement can be reached. Likewise, there should be independency between the classes i.e. allocation of requirements in to the classes of a structuring axes must not be based on the results of requirements structuring by another class [4].

It is also true to say here that consequently it is not possible to allocate requirements in one dimension but rather in an n-dimensional area, which should include all of the dimensions which need to be considered in the planning of an intra-logistical facility. The following nine dimensions apply for the model to be developed: legal aspects, environment, cost-efficiency, product, information, technical functional aspects, customer satisfaction and assessment and/or degree of performance. Temporal dimensions must also be considered. Although these do not form any independent dimension comparable to other dimensions it should not be ignored, as all requirements always contain a temporal element. For example, they can be found in the phase of facility construction or in the phase of facility operation (cf. Figure 3).

![Figure 3: Model for structuring requirements for intra-logistical facilities](image)

All the requirements made on an intra-logistical facility can be captured, structured and displayed using these nine dimensions. To do so, the individual dimensions must be detailed and sub-divided into categories so that the requirements may be adequately allocated to them. This shall ensure that requirements which refer to different aspects are not allocated to the same category and/or dimension. It shall furthermore ensure that requirements which indicate different abstraction levels are accordingly allocated to their respective degree of abstraction. This is a significant factor for the implementation of the requirement after structuring, e.g. with the help of a QFD. There follows more detailed examples regarding the dimensions “facility itself” and “environment”.

4.1. Structuring of requirements in line with the dimension “facility itself”

One possibility of structuring the requirement results from the product term or product model based on the St. Galler concept. It is possible to sub-divide a product in the observed case of an intra-logistical facility into the aspects of product core, formal product and extended product.

The sub-group “product core” comprises requirements which refer to the product material, the processing of the product or the function of the product. This includes requirements which refer to individual components of the facility
and requirements which comprise collaboration between components in the whole facility. In an intra-logistical facility these are e.g. requirements which describe directly the material of the rolls used, or requirements which describe precisely the individual drives. Likewise, requirements of the product core contain statements e.g. on the required performance and travel velocity.

Requirements which go beyond the product core are captured under the aspect of “formal product” whereby it is possible to make a further and finer sub-division in product design, packaging and brand name. Requirements e.g. concerning the colour of the facility, the facility (or component) manufacturer’s brand name and the ergonomics of the individual working place within the intra-logistical facility are also conceivable within an intra-logistical facility under the sub-division “formal product”.

Requirements concerning services for the product or those that fall under the aspect of maintenance, delivery, guarantees or consultations are to be allocated to the sub-group “extended product”. In addition to requirements of maintenance in an intra-logistical facility (accessibility, maintenance intervals, supply of spare parts), this sub-group must also consider requirements regarding the delivery of the facility (date of delivery, scope of delivery, packaging etc.). The following Figure 4 shows the contents of the dimension “facility itself”.

![Figure 4: Structuring in line with dimension “facility itself”](image_url)

### 4.2. Structuring of requirements in line with dimension “Environment”

Requirements may also be structured according to the aspect of “facility environment” in which the categories resources, environment and safety can be differentiated. In this dimension, requirements are allocated if they do not refer directly to the span of the facility’s the environment.

The category “resources” comprises the material and human resources which are required to operate the facility. Human resources can be quantitative and qualitative i.e. they include requirements concerning the number of required personnel but also the qualifications held by the personnel. Requirements regarding material resources can be differentiated into types of resources e.g. water, power, gas, steam etc.

The category “environment” can be sub-divided into macro-environment and micro-environment according to whether the requirements refer only to the direct environment of the facility or are of a global nature. The former comprises requirements of a social, political or technological nature and therefore refers to factors which cannot be directly influenced by the company. Requirements of the micro-environment refer to the facility environment and included aspects such as required space, radiation, noise etc.

The third category of this dimension comprises requirements referring to the safety of the facility whereby the aspect of occupational safety is of foremost priority. Such requirements are aimed at the design of the working place
Structuring stakeholders’ requirements concerning intralogistic facilities in the context of QFD

and the design of the facility. Hazardous points must be marked and/or secured to mitigate the danger of accidents and injuries to personnel. The following diagram shows a summary of the dimension “environment”.

![Diagram of dimension “environment”]

The model presented here allows therefore requirements of intra-logistical facilities to be allocated into the area spanned by the nine dimensions. A sub-division of the dimensions into categories and sub-categories combines requirements which refer to the same area e.g. the facility itself or the facility environment. By observing individual dimensions and/or selected categories of the dimensions, a thematic focus may be placed on certain groups of requirements. That is particularly important not only for implementing requirements into possible solutions but also for focusing on the needs of individual stakeholders, for example when looking at maintenance requirements.

5. SUMMARY

In the course of a research project, a model for structuring requirements made on intra-logistical facilities was developed based on an analysis of existing structuring strategies. This multi-dimensional model comprises nine dimensions and allows all requirements on an intra-logistical facility to be allocated to an area spanned by the dimension. The dimensions presented here must be further detailed as shown in the examples of “facility itself” and “environment” and is the subject of current research activities.

The authors wish to thank the Deutsche Forschungsgemeinschaft (DFG) for its support in this work.

REFERENCES


[12] Sonderforschungsbereich 696: Forderungsgerechte Auslegung von intralogistischen Systemen – Logistics on Demand; supported by the Deutsche Forschungsgemeinschaft (DFG) since 2006. www.sfb-696.de
Implementation of a Methodology for the Management of Corrective Actions through Quality Improvement Teams at a Manufacturing Company

Joao C.E. Ferreira and Kleber D. Ristof

1Universidade Federal de Santa Catarina
Departamento de Engenharia Mecânica, GRIMA/GRUCON, Caixa Postal 476
Florianópolis, SC, 88040-900, Brazil

ABSTRACT

In the current manufacturing scenario, where companies consistently seek to reduce waste in the production processes, and to improve their products, the correction of defects and the elimination of the basic causes of nonconformities have become essential activities for the success of the organizations. Consequently, it is important to implement effective processes of corrective action, which assure that actions are executed on the problems that have greater effect on the processes and products. This context led to the development of this work, which presents a methodology to carry out the analysis, prioritization and correction of nonconformities, having been implemented in a company that manufactures mechanization systems for agriculture. The developed work is based on multifunctional teams for the analysis, prioritization and execution of corrective actions on the identified nonconformities in different processes of the company, considering the following criteria: (a) rework costs; (b) recurrences; (c) effect of nonconformities on the process the products. Thus, different multifunctional teams were created to analyze and prioritize the identified nonconformities in different processes, and the nonconformities are analyzed by a committee that determines the actions to be executed on each of the prioritized nonconformities, besides forming a multifunctional group to act on the corrective action. After prioritizing the nonconformities, the whole process of corrective action is accomplished in the company’s corporate system of corrective action, using the 8D methodology for problem resolution. Among the results obtained with the implementation of this methodology it is pointed out the amount of executed corrective actions and those that are underway. Besides, it is also pointed out the dissemination of the corrective action process, where it is evidenced the importance of an effective analysis of the causes for the nonconformities in order to eliminate them, registering the actions in a system where all can visualize and track them.

1. INTRODUCTION

Many tools for the analysis and solution of nonconformities are known, and tools for prioritization are also available, so that problems of greater effect are treated in a different way in relation to problems of lower effect. However, in some cases, despite the availability of tools, they cannot be used in an effective way due to some particular reasons for each case. For instance, the lack of defined criteria for the prioritization of the nonconformities, or the excess of criteria that hinder the task of selecting the main problems, being extremely important the opinion of those who have direct contact with the problem to analyze and complete about its effect on the process as a whole. Based on that, the definition of a method for managing corrective actions, using appropriate multifunctional teams, with knowledge about each of the analyzed nonconformities, becomes viable due to the results that it can bring for the improvement and reduction of the problems that affect the manufacturing system.

The last step of the quality control process is the execution of actions to correct the deviations in relation to the objectives [1]. Methods and tools for the solution of problems are commonly used to gather information on a certain problem, to investigate their causes, and to propose an action plan to eliminate it and to prevent its recurrence. The intensity of the executed actions is directly proportional to the intensity of the problem, in other words, when a new problem occurs, immediate actions are executed to minimize it, but when its effect is controlled, the tendency is to attempt to solve a new problem, and in this way the first problem occurs again after some time [2]. Thus, the objective of the use of the methods and tools to solve problems consists of avoiding their recurrence, by applying all the steps recommended by the methods [2].

1.1 THE CONCEPT OF TEAMS AS A TOOL FOR QUALITY IMPROVEMENT

The greater the involvement of the organizations with quality, the companies realize the advantages of having people of all levels working together in teams, because many problems that need improvement will be
identified in processes and, from that, teams will be formed entrusted for finding solutions [3]. A single person using practices for quality improvement can contribute for a company to reach good results, however this person's knowledge and experience are not enough to understand everything that is involved in a process [3]. The work in teams can be an effective tool for total quality [4]. A team can be defined as "the result of the improvement of the relationship of a small group of people, when, in the presence of some special environmental conditions, certain operational cares are applied" [4].

1.1.1 QUALITY CONTROL CIRCLES (QCC)

The Quality Control Circles (QCCs) are "small groups formed by people of same areas who voluntarily develop activities to control quality" [5]. QCCs are formed by "voluntary employees who meet periodically to identify, to study and to improve work situations", and the conclusions of the studies are directed to the immediate manager under the form of suggestions to be implanted. The results obtained by QCCs found in the literature are surprising, leading to high productivity levels and quality through the collective effort [5].

1.2. 8D METHODOLOGY FOR SOLVING PROBLEMS

Ford Motor Company is responsible for the release of the MS 9000 manual (Materials Management System Requirement), which focuses on the procedure for the management of materials in the logistics sector of the organization, and the 8D methodology for solving problems in teams [6]. 8D is a methodology of solving problems that is significantly used in organizations, due to its simple and effective way to act on different types of irregularities. The 8D methodology assumes that a problem was already identified, and it consists of eight steps that are presented below [6]:

1. Establish the team
2. Describe the problem
3. Develop containment action
4. Define and verify the root cause
5. Select and verify possible corrective actions
6. Implement and validate corrective actions and verify the efficacy of the actions
7. Prevent recurrence
8. Wrap-up and recognize the team

2. STRUCTURE AND PRESENTATION OF THE DEVELOPED METHOD

The proposed method was implemented in a company that manufactures mechanization systems for agriculture. Some of the products in its portfolio are grain harvesters and seeders, and these agricultural machines are produced for internal and external consumption.

The QIT ("Quality Improvement Teams") method for managing corrective actions was applied to the considered company due to its need to have a consistent method of management of corrective actions for the identified nonconformities in manufacturing processes. Thus, a procedure was developed that defines the criticality of the nonconformities through multifunctional teams, and some criteria were applied for that. After the definition of the criticality of the nonconformities, those ranked as major or critical are taken to a committee of managers for the definition of the corrective action to be executed, the needed resources, and the period for concluding the actions.

If the nonconformities are classified as major or critical, they are registered in a software used by the company, where the records of corrective actions are also registered. The developed procedure uses this system to register the nonconformities and corrective actions. Thus, the definition of the resources to work on the solution of the problems and the establishment of the periods is made in this system, which is presented in greater details in this section. The procedure is composed of three stages, which are as follows:

(a) Information about the nonconformities

At this stage, the nonconformity is registered in the information management system, which is the system database. These pieces of information originate the reports that are analyzed later for the prioritization of the nonconformities.

(b) Priorization of the nonconformities

This stage includes the pre-priorization activities and prioritization of the nonconformities. The pre-priorization phase happens when a representative of the Quality department ranks the nonconformities, considering specific criteria for each one of the processes, such as: (a) manufacturing cost of the item; (b) number of occurrences of the same nonconformity; and (c) amount of nonconforming items. After
accomplishing the ranking of the nonconformities, they are taken to the analysis of multifunctional teams which, based on other criteria, classify the criticality of the nonconformities in critical, major, minor, or incidental. When a nonconformity is classified with a major or critical criticality, it is registered in the company’s system of registration of nonconformities and corrective actions. This system is called NCCA (“Non Conformance Corrective Action”).

(c) Definition of the corrective actions

All the nonconformities for which records are created in the NCCA system, are taken for a periodic meeting of managers or their representative individuals, called IRB (Internal Review Board) meeting. Thus, a representative of Quality has the responsibility of taking those registrations for the IRB meeting, which defines what action will be executed on the presented nonconformities. In this way, the actions can be executed through CAR or Issue. A CAR (Corrective Action Required) requires the use of the 8D methodology for executing the corrective action. On the other hand, an Issue does not require the use of the eight stages of the 8D methodology, since in this case the causes of the problem are already known. The solution of the problem is found by the team defined by the managers, who indicate those individuals that can contribute to eliminate the causes and to implement a definitive solution to the problem. Besides, people can be added to this team, if the leader of the corrective action considers it necessary. Figure 1 presents a flowchart of the method, which illustrates the activities to be carried out.

2.1 CRITERIA CONSIDERED FOR THE PRIORIZATION OF THE NONCONFORMITIES

Table 1 presents a summary of the activities considered in the analysis of the nonconformities through the Quality Improvement Teams (QITs), including their functions and some specifics for each of the analyzed processes.

2.2 PRIORIZATION MEETINGS OF THE QITS

2.2.1 OBJECTIVES OF THE QIT MEETINGS

The objective of the QITs meetings is to discuss on each of the pre-priorized nonconformities taken to the meetings, seeking to classify the criticality of each nonconformity. During the meeting, the group takes into consideration the criteria for the execution of corrective action (frequency, scrap/rework or adjustment/function, effect on operations), and in the end the multifunctional team defines the criticality of the nonconformities as critical, major, minor or incidental. Besides defining the criticality of the nonconformities, the team also discusses about the nonconformities classified as critical or major, so that more information can be added when the nonconformity is registered in the NCCA system. A facilitator, representing the Department of Quality, is responsible for conducting the QITs meetings.

2.2.2 INPUTS AND OUTPUTS OF THE QIT MEETINGS

The input to the meetings of each QIT consists of a spreadsheet with the ordered or pre-priorized nonconformities, according to the criteria presented for each of the processes. Then, taking into account the criteria for the corrective action, the group discusses about each of the ordered nonconformities in the spreadsheet, attributing a value for each of the criteria. The classification of the criticality of the analyzed nonconformities is based on the multiplication of the values defined for each criterion. When a nonconformity is classified as minor or incidental, it is considered again in the pre-priorization taken to the next meeting, and it may be analyzed again by the QIT, if its date of occurrence is within the considered time interval. Regarding the nonconformities classified as major or critical, they are registered in the NCCA system so that they are taken to the meeting of the managers committee for the definition of the action to be executed on the problem.

2.2.3 FREQUENCY OF THE QIT MEETINGS

The meetings of the QITs take place every two weeks. The period for each meeting is one hour, and thus this the groups do not have a specific number of nonconformities to evaluate, i.e. the nonconformities are analyzed and classified until the established duration of the meeting is completed. This procedure is applied to all the teams, except the process of Product Audit, which is performed weekly. In this latter case, as the audits are accomplished, meetings take place in order to classify the criticality of each of the identified nonconformities.
Figure 1: Flowchart of the QIT method

Table 1: General summary of the QIT method

<table>
<thead>
<tr>
<th>QIT</th>
<th>Function of the QIT</th>
<th>How the nonconformities arrive at the Quality Department</th>
<th>Pre-Priorization Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier</td>
<td>Determine the criticality of the nonconformities related to purchased items</td>
<td>Biweekly reports</td>
<td>Based on recurrence of the nonconformity; cost and quantity of the item</td>
</tr>
<tr>
<td>Assembly</td>
<td>Determine the criticality of the nonconformities identified in the assembly lines</td>
<td>Biweekly reports</td>
<td>Based on recurrence of the nonconformity and rework time</td>
</tr>
<tr>
<td>Welding</td>
<td>Determine the criticality of the nonconformities related to the welding processes</td>
<td>Biweekly reports</td>
<td>Based on recurrence of the same nonconformity</td>
</tr>
<tr>
<td>Primary Processes</td>
<td>Determine the criticality of the nonconformities related to the primary processes</td>
<td>Biweekly reports</td>
<td>Based on recurrence of the same nonconformity and the cost of generated scrap</td>
</tr>
<tr>
<td>Painting</td>
<td>Determine the criticality of the nonconformities related to the painting processes</td>
<td>Biweekly reports</td>
<td>Based on recurrence of the same nonconformity</td>
</tr>
<tr>
<td>Product Audit</td>
<td>Determine the criticality of the nonconformities identified in the product audit processes</td>
<td>Reports of audits carried out</td>
<td>Priorization is carried out directly, without the pre-prioritization</td>
</tr>
<tr>
<td>FPY Warehouse</td>
<td>Determine the criticality of the nonconformities identified in the first 15 hours of equipment run</td>
<td>Biweekly reports on identified problems</td>
<td>Based on cost of warranty and quantity if items with the same nonconformity</td>
</tr>
</tbody>
</table>
2.3 “IRB” MEETINGS TO DEFINE THE CORRECTIVE ACTION

2.3.1 OBJECTIVES OF THE IRB MEETINGS

The IRB (Internal Review Board) meeting takes place periodically and it is multifunctional, and it is composed of managers or their representatives to define the actions to be executed on the nonconformities classified by QIT as major or critical. Besides, this committee defines the available resources (people) and the period to eliminate the causes of each nonconformity. A facilitator, representing the Department of Quality, is responsible for conducting the IRB meetings.

2.3.2 INPUTS AND OUTPUTS OF THE IRB MEETINGS

The inputs to this meeting are all nonconformities classified with major or critical severity by the QIT, for which registrations are made in the NCCA system. The outputs are decisions regarding the opening or not of corrective actions for the presented nonconformities. In this way, if the decision corresponds to the opening of a corrective action, it is open through CAR or Issue in the NCCA system. On the other hand, the IRB may decide not to execute the action on the nonconformities, if such a corrective action is considered unnecessary. The option for not executing an action happens when the problem is a specific case or some deviation of the process, for which it is not still viable to allocate resources to work on the solution of the problem. If the decision is for not to execute the corrective action on the nonconformity, if such nonconformity occurs again, it is pre-prioritized again, and if it is classified as major or critical, it is taken back to the IRB to decide for executing or not the corrective action.

2.3.3 FREQUENCY OF THE IRB MEETINGS

The meetings take place weekly. The objective is to define actions for all the nonconformities classified as major or critical by the QIT meeting. However, if some nonconformity is not analyzed, for instance due to lack of information, the nonconformity returns with priority to the next meeting, and the people responsible for obtaining more data about the nonconformity are indicated.

2.4 SYSTEM FOR REGISTERING THE NONCONFORMITIES AND CORRECTIVE ACTIONS (NCCA)

The method developed in this work uses a corporate software to register the nonconformities prioritized by the teams, and also to register the corrective actions executed on these nonconformities. This software was especially developed for the company where this work was developed, and it can be accessed at any unit of the company, and therefore all the employees can verify the actions that were already executed on problems that might have happened. Figure 2 shows some search fields in the system, and through them the user can retrieve the actions that were already executed on the identified nonconformities. Thus, by inputting the information in any of the fields and clicking "submit", the user will have access to all the existing records related to the input information.

Figure 2: Some search fields for registering nonconformities and corrective actions
Through such technique the time to reach the cause and the solution of a certain problem that has already happened in the past can be reduced, when the executed action is properly registered. This allows the manufacturing system to depend less on the experienced professional responsible to solve those problems. Besides, the assembly line remains stopped for less time when the problem affects its activities. Such system, called NCCA (Non Conformance Corrective Action), is a tool based on the web, which uses the 8D methodology for executing corrective actions. The objective of this system, besides allowing the registration of the corrective actions, is to keep a database that allows the exchange of information among system users. For instance, if for a certain nonconformity that occurred at a certain unit of the company, some corrective action was already taken, when the same nonconformity occurs at another unit, the same corrective action already taken may be applied to solve the problem. Besides, NCCA is an integrated system with the capacity to interact with suppliers, in other words, when a nonconformity is related to a certain supplier, it can have access the system to analyze the registration of that nonconformity and, later, to register the executed actions to eliminate the problem.

2.4.1 ISSUE

A nonconformity taken to the Issue status must be traced, and a resolution process composed of four stages is applied, which are as follows: (a) definition and identification of the problem; (b) definition of the responsibilities; (c) definition of the containment action for the problem; (d) implementation of the permanent action. An Issue is understood as a corrective action performed on a nonconformity in which the causes are already known, just needing to implement solutions to eliminate the causes, and so in this case the 8D methodology is not needed to solve the problem.

2.4.2 CAR – CORRECTIVE ACTION REQUIRED

When a nonconformity is taken to the CAR (Corrective Action Required) status, the 8D methodology is used to execute the corrective action. Thus, the CAR is open when a problem presents a major or critical threat to the processes or to the customer, and the possible causes of the problem are not known.

2.5 CRITERIA FOR DEFINING CRITICALITY (NCCA CRITERIA)

In order to define the criticality of a nonconformity as critical, major, minor or incidental, there are three criteria that should be followed regarding the selection of the corrective action. These criteria, which are presented below, are taken into consideration by the QITs for classifying the criticality of the analyzed nonconformities.

2.5.1 FREQUENCY CRITERION

The frequency of a nonconformity is determined by the number of times that the problem occurred, taking into consideration the source that caused the problem. Table 2 presents the values for the frequency criterion according to the source pointed out in the problem report, and also according to the number of occurrences of the same nonconformity in the processes considered.

<table>
<thead>
<tr>
<th>Frequency Value</th>
<th>Source and Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – First Occurrence on the Shop Floor</td>
<td>The problem is the first occurrence on the shop floor</td>
</tr>
<tr>
<td>2 – First Occurrence in Product Audit</td>
<td>The problem is input from a field or audit report</td>
</tr>
<tr>
<td>3 – Recurrence or Multiple Occurrence</td>
<td>The problem occurred in multiple occasions</td>
</tr>
<tr>
<td>4 – Recurrence in Product Audit</td>
<td>The problem is a recurrence of a product or field audit</td>
</tr>
<tr>
<td>5 – Multiple Recurrence</td>
<td>The problem occurred multiple times, independently of the source</td>
</tr>
</tbody>
</table>

2.5.2 SCRAP/REWORK OR ADJUSTMENT/FUNCTION CRITERION

This criterion takes into consideration two situations: (a) the cost of scrap or rework caused by the identified nonconformity; (b) the effect of the nonconformity on the product. In this case, the value of the worse effect is the one attributed to this criterion, as presented in table 3.

<table>
<thead>
<tr>
<th>Value</th>
<th>Values for Scrap/Rework or Adjust/Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Less than US 200.00 or a cosmetic defect in the product</td>
</tr>
<tr>
<td>2</td>
<td>US 200.00 to US 1,000.00 or some effect on the adjustment of the product</td>
</tr>
<tr>
<td>3</td>
<td>US 1,000.00 to US 2,500.00 or adjustment suffers a significant effect</td>
</tr>
<tr>
<td>4</td>
<td>US 2,500.00 to US 5,000.00 or some effect on the function of the product</td>
</tr>
<tr>
<td>5</td>
<td>More than US 5,000.00 or a highly negative effect on the function or on the reliability of the product</td>
</tr>
</tbody>
</table>
For instance, if the cost of scrap or rework caused by a certain nonconformity is lower than US$ 200.00 (value equal to 1), but the adjustment of the product is significantly affected (value equal to 3), then the higher number (i.e. 3) is the value attributed to this criterion.

2.5.3 EFFECT ON OPERATIONS CRITERION

The third criterion, called Effect on Operations, is determined by the effect of the nonconformity on the operations as a whole. Table 4 presents the values for this criterion.

Table 4: Values for the criterion Effect on Operations

<table>
<thead>
<tr>
<th>Value</th>
<th>Effect on Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimum effect on the flow of operations</td>
</tr>
<tr>
<td>2</td>
<td>Delays exceed takt time of the assembly line</td>
</tr>
<tr>
<td>3</td>
<td>Item withdrawn from production or requires repair outside the assembly line</td>
</tr>
<tr>
<td>4</td>
<td>Delivery of the product is delayed</td>
</tr>
<tr>
<td>5</td>
<td>Delivery of the product is retained; assembly line is stopped; or shipping is stopped</td>
</tr>
</tbody>
</table>

After defining the values (1, 2, 3, 4 or 5) for each of the criteria, the three values are multiplied, and the result of the multiplication leads to the determination of the severity or criticality of the nonconformity. Table 5 presents the classification factors based on the results of the multiplication of the values attributed to the criteria.

Table 5: Categories of Criticality based on the Classification Factor

<table>
<thead>
<tr>
<th>Classification Factor</th>
<th>Categories of Criticality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification Factor &lt; 8</td>
<td>Incidental</td>
</tr>
<tr>
<td>8 ≤ Classification Factor &lt; 27</td>
<td>Minor</td>
</tr>
<tr>
<td>27 ≤ Classification Factor &lt; 64</td>
<td>Major</td>
</tr>
<tr>
<td>Classification Factor ≥ 64</td>
<td>Critical</td>
</tr>
</tbody>
</table>

If any of the criteria for classifying the severity of the nonconformities (i.e. Frequency; Scrap/Rework or Adjustment/Function; Effect in the Operations) is attributed a value equal to 5, the severity of the nonconformity is automatically classified as belonging to, at least, a Major category.

3. RESULTS

The general results obtained regarding the number of corrective actions open until now are presented in table 6, whereas tables 7 and 8 show the results related to purchased and manufactured items, respectively.

Table 6: General results of the QIT and the IRB meetings

| QIT / IRB – Results obtained (until Jan/20/2009) |
|-------------------------------------------------|-------------------------------------------------|
| Total of open nonconformities by the QITs        | 238                                             |
| Pending decisions of the IRB                     | 0                                               |
| Taken decisions of the IRB                       | 238                                             |
| Open CAR                                         | 35                                              |
| Open ISSUE                                       | 105                                             |
| Decision for not executing corrective action     | 94                                              |

Table 7: General results of the QITs and the IRB meetings: purchased items

| QIT / IRB – Results related to the purchased items (until Jan/20/2009) |
|------------------------------------------------------------------------|-------------------------------------------------|
| Total of open nonconformities by the QITs                              | 171                                             |
| Pending decisions of the IRB                                          | 1                                               |
| Taken decisions of the IRB                                            | 170                                             |
| Open CAR                                                              | 32                                              |
| Open ISSUE                                                            | 68                                              |
| Decision for not executing corrective action                          | 68                                              |

Table 8: General results of the QITs and the IRB meetings: manufactured items

| QIT / IRB – Results related to the manufactured items (until Jan/20/2009) |
|--------------------------------------------------------------------------|-------------------------------------------------|
| Total of open nonconformities by the QITs                                | 67                                              |
| Pending decisions of the IRB                                            | 1                                               |
| Taken decisions of the IRB                                              | 66                                              |
| Open CAR                                                                | 3                                               |
| Open ISSUE                                                              | 37                                              |
| Decision for not executing corrective action                            | 26                                              |
3.1 EXPECTED RESULTS IN A MEDIUM TO LONG RANGE

With the increased experience of using the proposed method, some expected benefits of its application in the medium and long term can be pointed out, which are as follows: (a) reduction of rework costs and repairs; (b) reduction of warranties costs; (c) reduction in the amount of equipment failures; (d) increase of production capacity; (e) reduction in the amount of nonconformities; (f) improvement of product quality; (g) increased profits.

4. CONCLUSIONS

The implementation of the method for managing the corrective actions was very important, because it identifies a significant amount of improvement opportunities. In the past, in the company where this work was implemented, a reduced number of corrective actions were registered in the corporate system, but in fact a large number of problems existed, which were controlled through spreadsheets. Thus, the parallel activities for controlling the corrective actions are being eliminated, and the NCCA system remains as the only control method. Also, the method also helped improve the quality of information for opening a corrective action, which facilitates the process of analysis of the root cause, reducing the time for closing the corrective action.

Considering the results obtained until now, related to the opening of corrective actions, since the method is relatively new, we consider that if the developed method continues to be conducted in an effective way, it can lead to good results in the medium and long term for the company. By observing the presented numbers, it can be noticed a significant difference among the amount of open corrective actions for purchased items in relation to the manufactured ones, which is due to the expressive amount of items outsourced by the company recently.

REFERENCES

Improving transport productivity within a food distribution SME through a simplified six sigma methodology

Alireza Shokri*, Farhad Nabhani, Nazir Hossien
School of Science & Technology, University of Teesside
Tees Valley, Cleveland, TS1 5JU, UK

ABSTRACT
The purpose of this paper is to promote the utilization of a problem solving methodology to improve transport productivity of a food distribution in small to medium enterprises (SME). A simplified version of a six sigma methodology was adopted to identify the defect, root causes of the defect and improve two major transport key performance indicators (KPIs). It was found that there are some internal optimum solutions which directly target the causes of the most critical defects in each key performance indicator (KPI) for a food distribution SME. This has potential impacts on the cost efficiency of the transportation. This research has highlighted only two KPIs due to limited time. But, there are other key measures in the same environment which can be analysed. There was also a challenge to find Six Sigma sources with the experience in logistics and transport productivity as the potential respondent. Implementing a systematic problem solving methodology could identify the potential and actual causes of the defects and variables in transport section of food service addressing lean principles, customer satisfaction and profitability. This paper has focused on specific areas of food transport productivity for a food distribution while utilizing the simple set of tools and technologies of the Six Sigma to find the root causes of the defects.

1- Introduction

Supply Chain Management (SCM) is one of the key effective factors of the competitiveness and business improvement in any industry. There are numerous statements from researchers associated with the definition of SCM [1][2]. SCM includes the management of information systems, purchasing, customer service, sourcing, transportation, scheduling, order processing, and inventory management, warehousing and marketing. The success of SCM between or within firms depends upon the continuous performance measurement of the Supply Chain through quality initiatives or benchmarking. The importance of Supply Chain was emphasised by several scholars [3 - 6]. Only a few studies [7] have analysed the management of Supply Chain performance [6][8].

The application of quality initiatives in Supply Chain which are based on performance measurement was also highlighted by researchers [9 - 15]. Logistics Management is a key element of Supply Chain which was analysed more frequently by the researchers [6][15 - 19]. Logistics Management focuses on internal activities in Supply Chain operation. There has always been confusion between logistics management and Supply Chain management. This was highlighted by some researchers [2].

Logistics operation includes purchasing, distribution, the inventory management, packaging, manufacturing and customer service [20]. Logistics is generally viewed as within one company although it manages flow of goods between the companies for the purpose of confirming to customer requirements. Therefore, SCM includes the logistical flows. Chiu (1995) indicated four different flows in logistical system [16]:

- Material Flow
- Merchandise Flow
- Money Flow
- Information Flow

It appears that all these four flows are related to each other and any failure or success in one flow will have dramatic impacts on the performance of the other flows. Managing these flows is literally defined as the logistics management [16].

There are different researchers that emphasised the critical role of logistics management on flow of material, money and information, business improvement, and competitiveness [17][20 - 22]. It is also suggested that logistics management is vital for retailing, transport, distribution or any other service orientated industry [16][17].

* Corresponding Author: Tel: (+44)1642252121, Fax: (+44) 1642252821, E-mail: Shahab@dlsne.co.uk
Transport management is an element of logistics management that have direct impacts on material flow and physical distribution. In fact, the literature suggested the Transport and Distribution are the major physical activities in logistics service providers [17]. Transport productivity is a quality measure in logistics management which underlines the importance of the cost and service efficiency. It means improving transport productivity via reducing the cost or improving the distribution service will improve both the profitability and competitiveness in supply chain. It appears that transport productivity has become a determining measure for every industry to be economically cautious in existing market.

There are different literatures highlighting the effect of adverse market condition, volatility in oil prices and tightening credit market on general and logistics industry in particular [18]. This has instigated a crisis management need in every industry to control the costs.

It is the intention of this paper to investigate challenges of the cost of poor quality in transportation. The paper aims to provide a systematic methodology of quality management initiative to identify and solve the critical problems associated with transport productivity of food distribution in Small – Medium Sized Enterprise (SME). The research program analysed two distinct key performance indicators in Transport and Distribution through two different cases studies. The results suggested the optimum solutions in these two KPIs in order to increase the productivity and competitiveness.

2- Background for the research

There are so many critical factors in transport productivity which were studied by the literatures. Shorter cycle time, higher frequency of delivery, lower delivery delay rate and transportation cost are within these important elements. There fore, a deeper literature review was undertaken in transportation cost and lead time in logistics. It was crucial to have more in – depth research in logistics management in order to capture more valuable support for the real world case study in this research program. As a result, the process of “Order – to – Delivery” (OTD) was adopted, since this process formulates the important part of strategic and operational decisions in improving the transport productivity for a Logistics SME. Mattsson, 2004 and Forslund, 2009 suggested OTD process as one of the most important processes in logistics which includes sub processes as [6][23]:

- Customer’s ordering
- Supplier’s delivery
- Logistics Service provider’s (LSP’s) transportation
- Customer goods receipt

A further literature review was undertaken in “Transportation” sub – process to identify the key performance indicators (KPIs) for this sub-process. Matsson, 2000 and Forslund, 2009 defined the transportation as an integrated cross-company process which is considered as the fulfilment and procurement process together with the company individual process [24][6]. Forslund, 2009 has also highlighted the importance of transportation in quality of OTD process [6]. Gunasekaran, 2003 has proposed a conceptual strategic model in successful logistics and transport management after identifying the major problems and critical success factors in logistics. Good planning of logistic systems was one of the key successful factors which its absence or failure will result in too many problems associated with transport efficiency and cost [17]. Accordingly, Gunasekaran, 2009 suggested that minimising fuel cost and optimising the routing and scheduling could be effective strategies in transpiration and logistics planning [17]. Mckinnon, 1999 had before suggested that Fuel efficiency and deviation from schedule are among the KPIs in a Food Logistics and transportation [25]. Therefore it was decided to have more focused literature review on fuel efficiency and deviation from schedule, since they are strategically associated with transport productivity.

There are different literatures in Transport lead time and cycle time to indicate the effects of lead time on transport efficiency and customer satisfaction. [6][17][19][26-27]. Lead Time is considered as waste in any service orientated organisation [28 - 30] and is an important customer satisfaction measure in transportation process.
The extensive amount of research is concerned with the reduction of lead time \cite{6,8,31-34}. It was even highlighted more by Keebler et al, 1999 and Forslund 2007 that on-time delivery and lead time reduction are two of the most important dimensions in OTD process \cite{8,33}.

Forslund, 2009 offered different lead time dimensions such as lead time accuracy, lead time efficiency, lead time variability and lead time adaptability \cite{6}. He particularly suggested that these dimensions are critical in satisfying the customer, while they are quality characteristics of the transport process in a logistics organisation. Therefore, any problem in these dimensions which affects the customer satisfaction can be identified as the defect in quality improvement term. “Deviation from schedule” has dramatic impact on lead time adaptability and lead time accuracy in any OTD process. This study considered the “deviation from schedule” as an important defect and it was decided to deal with this problem through a quality improvement methodology in the real world case study.

It is understood that so many studies have reported the cost of physical distribution in logistics \cite{16,35-37}. They all agreed that fuel efficiency has an important role in planning efficiency and physical distribution logistics. The fuel efficiency forms the major block of the physical distribution cost. Having identified the transportation cost as one of the key measures in transport and logistics, fuel efficiency is a crucial measure to be analysed through a systematic quality improvement methodology. The importance of implementing quality management programs in transportation has already been highlighted by the literature \cite{38}.

Six Sigma is a business strategy which embraces the quality improvement methodology and is one of the most reliable problem solving \cite{39,40} which could be applied in transportation. The application of Six Sigma in Supply Chain and logistics in SMEs has already been promoted by different studies \cite{39-42}. Six Sigma is a disciplined method of using rigorous data gathering and statistical analysis to pinpoint sources of errors and ways of eliminating them \cite{43}. It is also suggested that it can be applied in more simplified and less complicated methodology which suits SME businesses \cite{39,44-45}. Therefore, it was decided to select two distinct Six Sigma based case studies associated with two KPIs in order to identify the defects, detect the causes of the defects and suggest the solutions. This would help to increase the reliability of the suggested potential solutions, since the study was conducted under a systematic data driven methodology.

3- Research Methodology

Applying a step-by-step process based road map is a key success factor in implementing any Six Sigma project regardless of the size or type of the business. The Six Sigma methodology that was used in this case study was DMAIC (Define, Measure, Analyse, Improve, Control). This methodology is the most common reactive methodology in Six Sigma to solve the problem. It leads the improvement process through laser focusing and helps detecting the root causes of the failure in any process \cite{40,46}. Hence, this methodology was used to study the process of transport in a food distribution SME. It was decided to focus on each KPI separately and apply the methodology to identify the defects, collect the data, find the root causes of the problem and suggest the optimum solution. The key contents of this methodology are the tools and technologies that are used. There are different functional tools and techniques in each stage of DMAIC and it was intended to use more straight forward tools and techniques in this case study due to limited resources in SME. Table 1 shows the practiced tools and techniques for this case study.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Tools &amp; Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Define (D)</td>
<td>Pareto Chart, Histogram, Quality Function Deployment, Brainstorming, Questionnaire</td>
</tr>
<tr>
<td>Measure (M)</td>
<td>Data Collection, SIPOC, Process Sigma calculator, Measurement system analysis</td>
</tr>
<tr>
<td>Analyse (A)</td>
<td>Cause &amp; Effect XY Matrix, Brain Storming</td>
</tr>
<tr>
<td>Improve (I)</td>
<td>Affinity Diagram, Brain Storming, Analytical Hierarch Process</td>
</tr>
<tr>
<td>Control (C)</td>
<td>Monitoring Chart, Process Sigma calculator</td>
</tr>
</tbody>
</table>

The methodology of DMAIC was initially applied for the first KPI which was “Deviation from schedule”. This metric was selected since it reflects the level of customer service to meet the scheduled time in delivery.
The delivery operation of the company was closely observed and the data were collected through data base for four weeks. It was agreed that standard deviation in each route would determine the performance of each route associated with “Deviation from Schedule”. The collected data were the spent time on and between delivery spots in each delivery route to calculate the average time and standard deviation. The standard deviation for each run was calculated. Table 2 represents the runs with their standard deviation in two occasions of delivery.

Table 2- Average Standard Deviation for each Route

<table>
<thead>
<tr>
<th>Route No</th>
<th>on spot time spent</th>
<th>Traffic (Shop to shop)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.498</td>
<td>1.369</td>
</tr>
<tr>
<td>2</td>
<td>3.8435</td>
<td>2.3311</td>
</tr>
<tr>
<td>3</td>
<td>3.942</td>
<td>2.358</td>
</tr>
<tr>
<td>4</td>
<td>10.33</td>
<td>23.625</td>
</tr>
<tr>
<td>5</td>
<td>11.86</td>
<td>6.026</td>
</tr>
<tr>
<td>6</td>
<td>2.23</td>
<td>3.81</td>
</tr>
</tbody>
</table>

Then, the team decided to identify the most defective routes through counting the number of runs in each route with higher standard deviation than average. Figure 1 graphically indicates the three most defective routes. It was revealed that routes 3, 4 and 5 were the most defective routes to be focused more in further steps.

The Voice of the Customer (VOC) in the “Delivery” process was collected through interview with the customers and their requirements were applied in Quality Function Deployment (QFD) to introduce the most important customer requirement in the company as the Critical to Quality (CTQ). The specific scoring procedure was applied in QFD and the most important customer requirements were introduced as “On time delivery” and “Consistency in Delivery”. This would lead the team to laser focus on these two measures. The define stage was completed in this phase. The current performance of the delivery was measured in the measure stage. The defect rate was measured within the total number of runs of the whole routes for the period of four weeks. The defect rate was calculated as 38% representing 1.81 Sigma level.

Having identified the CTQ and the most defective routes, the root causes of the defect was identified through brainstorming and analysis in Cause & Effect XY Matrix. Table 3 represents the result of the analysis to find the potential causes of the defect. The scoring procedure was based on brainstorming with academics and other businesses, and interview with the operational team including drivers.
The data analysis found four potential causes as the most important sources of the defect to be highlighted in improvement stage. The brainstorming and interview were conducted concerning the eliminating or minimising the four important causes. Further analysis recommended that there were no special causes of the defect as the process was under control.

The potential solutions were discussed through indicating the header for the solutions. The best solution under each header was selected through brainstorming, interview and the relationship scores between “customer requirements” and “technical requirements” in QFD process. Table 4 represents these solutions.

Table 4 – The potential solutions for the 1st KPI

<table>
<thead>
<tr>
<th>Header</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Customizing route depending on opening hour (No 1)</td>
</tr>
<tr>
<td>Internal Operation</td>
<td>Driver Training (No 2)</td>
</tr>
<tr>
<td>Customer Relationship</td>
<td>Customer Training (No 3)</td>
</tr>
<tr>
<td>Resources</td>
<td>Satellite Navigator (No 4)</td>
</tr>
</tbody>
</table>

Then, the Analytical Hierarchy process (AHP) was applied to identify the most optimum solution to minimise or eliminate the causes of the defect. This would help management team to focus on single strategic decision making with the most effective role in operation. These four solutions were discussed carefully within the organisation through brainstorming, reliability and applicability review to select the most optimum solution to be implemented as figure 2 represents the analysis.
Figure 2 – AHP Analysis on potential solutions to select the most optimum solution for the 1st KPI

The AHP analysis indicated that “Customer Development” (No 3) is the most optimum solution (0.459) to reduce the causes of the “Deviation from Schedule”. This measure involves customer training to how to deal with the incoming delivery and payment procedure. This means that company needs to develop the customer in order to satisfy them in better customer service.

The second KPI which was selected for this case study was “Fuel Efficiency”. It was decided to monitor the performance of the delivery vehicles closely in the company for four weeks in terms of “Fuel Efficiency”. The current level of fuel efficiency for the vehicles was calculated through below equation.

\[
Fe = \frac{\text{Lw/M}}{F \times C}
\]

Where LW = the average loaded weight on each Vehicle
M = average mileage for the individual Vehicle in each Run
F = Average fuel fill for each Vehicle
C = Average fuel cost in pence

Figure 3 depicts that Vehicles D, F and G have the worst fuel efficiency records among the delivery vehicles. This will help the team to be more accurate in identifying the root causes of the problem.

As the key step in “Define” stage, the VOC in Delivery process for this KPI was collected through interview with the “Service and maintenance” providers for the vehicles and delivery team. The customer requirements were applied in QFD process to introduce the CTQ. The specific scoring procedure was applied in QFD and the most important customer requirements were introduced as “Less fuel consumption” and “More productive delivery”. The define stage was completed in this phase. The current performance of the delivery was measured in the measure stage. The number of runs for each vehicle with less than “average fuel efficiency index” was indicated as the defect to calculate the total number of defects for the whole number of runs carried out by the vehicles in four weeks. The defect percentage was 31% representing 2 Sigma level.

Having identified the CTQ and the most defective vehicles, the root causes of the defect was identified through brainstorming and analysis in Cause & Effect XY Matrix. Three key variables affecting fuel efficiency were identified after reviewing the performance of the defective vehicles to be analysed by identifying the most related
Improving transport productivity within a food distribution SME through a simplified six sigma methodology

des are the causes of these variables. The analysis was carried out through brainstorming and data analysis of the vehicle records.

Table 5 – Cause & Effect XY Matrix to reflect the causes of the defects in second KPI

<table>
<thead>
<tr>
<th>Input/Process Variables (X’s)</th>
<th>High Fuel Cost</th>
<th>Poor Van Operation</th>
<th>Poor Driver Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Transport Planning</td>
<td>9</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>More fuel consumption</td>
<td>9</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Less drivers efficiency</td>
<td>3</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Less weight productivity</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>van safety</td>
<td>0</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Poor engine operation</td>
<td>9</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Poor route scheduling</td>
<td>9</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Too much travelling</td>
<td>9</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5 represents four potential causes as the most important sources of the defect to be highlighted in improvement stage. The brainstorming and interview were conducted concerning the eliminating or minimising the four important causes. Further analysis recommended that there were no special causes of the defect as the process was under control.

The potential solutions were discussed through indicating the header for the solutions. The best solution under each header was selected through brainstorming, interview and the relationship scores between “customer requirements” and “technical requirements” in QFD process. Table 6 represents these solutions.

Table 6 – The potential solutions for the 2nd KPI

<table>
<thead>
<tr>
<th>Header</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel</td>
<td>Cost effective fuel brand (No 1)</td>
</tr>
<tr>
<td>Operation Management</td>
<td>Regular and reliable servicing by the service provider (No 2)</td>
</tr>
<tr>
<td>Customer Development</td>
<td>Rout standardisation with customer confirmation (No 3)</td>
</tr>
</tbody>
</table>

Then, all three solutions were analysed through AHP to select the most optimum solution which lead the minimising or eliminating the root causes of the defects. This would help to implement the most reliable and applicable solution. Figure 4 shows the AHP analysis.

![Figure 4 - AHP Analysis on potential solutions to select the most optimum solution for the 2nd KPI](image-url)
4- Control of the implemented solutions and Discussion

The optimum solutions for each KPI were implemented for the period of four weeks to monitor the performance of the transport process in respect to minimising the actual causes of the defect. In respect to reducing the deviation from schedule, some customers were selected randomly in each business category and the training was conducted through information sharing and in house supervision to identify the weaknesses in delivery point. The customers were also asked to arrange the appropriate and standard opening hour to help the transport and shop floor team to organise their delivery within the route. Having implemented these strategies, the new set of data was collected. Table 7 represents the comparison of the performance level before and after implementing the solutions. The optimum solutions for improving the fuel efficiency were also implemented by the transport team. It was decided to arrange a special team to investigate the fuel supply options and monitor the regular service and maintaining jobs from service providers. More intensive fuel market research, establishing a fuel branding policy in the company, re-scheduling of the routes and increasing the sales opportunity were the implemented strategies to increase the fuel efficiency in the company. The new set of data was collected after implementing these strategies for four weeks. Table 7 also represents the comparison of the performance level before and after implementing the solutions.

Table 7 - Comparison of the KPI performances before and after the analysis

<table>
<thead>
<tr>
<th>KPI</th>
<th>Before Implementation</th>
<th>After Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Defect Percentage</td>
<td>Sigma Level</td>
</tr>
<tr>
<td>1st KPI</td>
<td>38% 1.81</td>
<td>23% 2.24</td>
</tr>
<tr>
<td>2nd KPI</td>
<td>31% 2</td>
<td>14% 2.6</td>
</tr>
</tbody>
</table>

Table 7 indicates that the effect of DMAIC methodology on the performance of the KPIs was significant. The defect percentage was reduced in both KPIs, whilst the Sigma level increased. This means that the quality of service and cost efficiency for the sub-process of transport in the company has improved through implementing the DMAIC methodology in logistics process of a food distribution SME.

5- Conclusion

It is concluded that quality improvement programs can have dramatic roles in improving the logistics performance. Six Sigma methodologies have a systematic problem solving approach in targeting the critical elements in logistics process and improve their quality. DMAIC methodology could dig out the real sources of any problem in logistics and promote more productive and profitable operations without requiring any complicated tool or technology. It means, simplified version of DMAIC methodology could easily be utilized in logistics organisations to improve the level of customer satisfaction and profitability. This methodology could be applied for any process or sub-process in logistics and supply chain to improve the quality. There are so many other KPIs in logistics operation that can be approached in further research studies through the same methodology.

References

Improving transport productivity within a food distribution SME through a simplified six sigma methodology


Requirements for Roll Conveyors and their (Inter-)Dependencies
Based on Domain Knowledge

Horst-Artur Crostack*, Jonas Mathis, Robert Refflinghaus
Chair for Quality Engineering
Dortmund University of Technology
44227 Dortmund, Germany

Abstract
When planning and designing material flow facilities according to customer needs, a large number of requirements and their impact on each other have to be taken into account. In the German Collaborative Research Centre 696, those dependencies have been investigated using the example of roll conveyors. The research objective is to analyse different kinds of dependencies between the physical and logical parts of an intra-logistics facility. Abstract concepts used in the formulation of requirements have been subdivided into more useable ones, e.g. the notion of "flexibility" has been given various different interpretations depending on its context. Moreover, several known approaches have been combined and adapted to the domain of intra-logistics to form a new model for technical, functional and logical interdependencies.

1. Introduction

Today, it is of prime importance to ensure that goods are transported in the best time-efficient and cost-effective way. In addition to the macroscopic or global aspect focus is now being increasingly placed on the procedures and organization of in-company goods conveyance: the sector of intra-logistics which deals with the conception, use and control of material flow systems for in-house logistics [1]. Seen from the quality management point of view, intra-logistics is an interesting sector of application as the many stakeholders along the life cycle of a facility demand requirements of the system with varying degrees of abstraction and with diverging or contrary objectives. On top of that, the whole life cycle of the facility is determined by the requirements of the intra-logistics stakeholders (manufacturer, operator, state, maintenance, order pickers on the conveyor etc.) even if these can only be taken into account at the beginning of the product life cycle in facility development. Hence, bad planning can result in the need for a high level of technical expertise and especially in costs (“Rule of Ten”). Facility planning directed more strongly towards customer requirements results in cost and productivity advantages for both manufacturer and operator alike as up to now, for example, tendencies to oversize the facility and neglect the service side have been observed [2]. Likewise with reference to reduction on design changes and avoidance of inaccurate or incorrectly oriented facility concepts, early consideration of requirements in an integrated system constitutes a very major element. Optimal facility design must therefore be supported with additional tools. The numerous requirements also mean that computer-assisted processing of the requirements is indispensible. Hence, processes for EDP-assisted facility planning are being investigated as part of the 696 research field [3]; processes which take into consideration the many requirements and the diversity of the specialist domains.

1.1. Objectives of the German Collaborative Research Centre 696 and this paper

Up to now, the planning and the design of intralogistics facilities is an engineering-driven process mainly based upon the designer’s previous experience with similar cases and the respective requirements [1, 2, 4]. Several structured techniques such as “Logistics Function Deployment” [4] have been proposed, neither of which could be established on a broader scale.

The collaborative research centre 696 is occupied with central issues concerning requirement-oriented planning of intra-logistic systems and their sub-components such as new drives, optimum roller spacing, etc. The sub-project D –
the findings of which are partly reported here – is focused on methodical and data-technical handling of individual requirements and aims at developing concepts to implement these later using certain methods of optimisation [2]. An important point is that these methods have to support the understanding of the terminology and language of the stakeholders involved. The sub-objectives of the project for an initial requirement processing system include the structured storing of stakeholder requirements in a data model (“template”), the provision of a general domain model which displays expertise from the application domain of intra-logistics and thus a “common language” for the stakeholder and a “semantics” (context) for computer-assisted requirement analyses and methods for a in-depth requirements analysis, whereby requirements are referenced with or related to one another.

These are sub-steps along the way to customer-directed facility planning; i.e. a standardized and complete capturing of requirements, some of which will be shown as examples in the course of this paper using the roll conveyor as a basis.

1.2. THE APPROACH OF THE SUB-PROJECT AND RELATED WORK

This paper addresses primarily some of the methods used in requirements analyses in particular the relations model used for describing the inter-activities occurring between requirements. This model is strongly based however on the data modelling for requirements used in the sub-project. It images requirements in a “template” which can be compared to a file card in order to capture all of the information contained in the requirements in a structured way and as completely as possible. Sections 2 and 3 of this paper will therefore be looking at the data model as a basis for further investigations. The following sections then address the various types of relations and inter-activities which describe the influence requirements have on one another. To do so, existing expertise is deposited in the domain model, although this module which deals with knowledge management and in some parts with language technology will not be looked at in any great detail. The examples have been selected in such a way that they can be comprehended without any explicit references having to be made to domain knowledge.

In creating the requirements template and the analysis of requirements relations, German-language papers on construction methods (e.g. [5]) for machine and plant manufacture ([6-8]), in particular from the automotive construction sector [9-11], systems and requirements engineering [12, 13], papers handling the research fields of knowledge management [14, 15] and language research [16] were also used as supportive literature.

One of the main tasks was to analyse, combine, refine and adapt a selection of these known approaches whereby, due to the strong current interest in intra-logistics as a field of research and efforts to support “intelligent” processes in facility planning, the project work presented here has become of prime importance. Moreover, semantic technology [17] is used to form the basis of understanding and comparing stakeholder. Refer to [9] and [14] regarding preparatory work in this field.

1.3. INTRA-LOGISTICS AND ITS FIELDS OF APPLICATION

It was determined that some sort of material flow is present in practically every company in the manufacturing sector. Material flow systems are employed for various purposes:

- the product which is to be manufactured is passed through separate physical production steps
- the pre-products are brought together and/or led to their processing site
- different products are brought into and out of a storage system using a material flow system. Additionally, goods order picking operations are often supported (e.g. in the mail order sector). Figure 1 shows an example of a standard process for such a case.
- the material flow system itself can fulfil important business functions such as the conveying, sorting and diverting of letters and parcels in the postal sector

![Figure 1: Schematic diagram of standard processes in warehouses as per [1]](image-url)
Intra-logistical facilities can also be found in filling units (bottling plants) or in waste sifting stations – a completely different setting with mostly different requirements. For example, the work in [1] presents a general view on structuring material flow components according to a variety of criteria (type of goods to be transported, continuity of transport, etc.). It emphasizes the diversity of components used within a facility which can forcibly limit the object under observation. Although we claim our observations to be generally valid regarding intra-logistical facilities on the whole, investigations were based on the roll conveyor as a typical example of a continuous means of transport, thus keeping the area of observation as comprehensible as possible. The roll conveyor can be used in a variety of setting and, in spite of its comprehensible functionality, comprises quite a high level of complex interactions which need to be taken into consideration when implementing the requirements.

2. DEVELOPMENT OF THE DATA MODEL

The starting point for all observations or requirements made of a roll conveyor is a suitable representation of the information needed. Requirements identified in interviews or which, for example, originate in safety prescriptions are available in written form and their contents need to be structured. The objectives of the project demanded that all information was to be able to be evaluated by people as well as by computers. The use of an intra-logistical domain model (terms and their coherency, [17]) adds context information to the data thus providing “semantics” for data input. Therefore, this project works with a data model, the structured “requirements template” and connects it closely with the domain model. This way, information can then be later transferred either by domain experts, or semi-automatically via the requirements analysis system. However, this means that requirements from very diverse sources, in a variety of “domain languages” and with differing depth of detail must be handled coherently. A possible spectrum of requirements might range from “The means of transport must be particularly energy-efficient.” or “The facility must be as flexible as possible.” to “The throughput of the roll conveyors used must be at least 50 parcels per hour.”
and lastly the property characteristics. Splitting the object up into an additional structure permits working with a vocabulary of concepts in the domain of “intra-logistics” (ontology). This idea is shown in Figure 3.

3. PROPERTIES IN REQUIREMENTS MODEL AS A BASIS FOR REQUIREMENT RELATIONS

The following section looks at the specialties of this data model with respect to the analysis of relations. Particular emphasis is placed on analyzing the degree of detail of the properties which are to be modelled.

3.1. ABSTRACTION AND LEVEL OF DETAIL

The addressed level of detail describes a very “soft” property of a requirement: how roughly or how exactly are the properties which must be met by the facility described by the (textual) requirement? As there is no pre-defined measure for this “accuracy” there is no fixed ranking i.e. scale for the depth of detail of a requirement, especially considering the variety of requirements occurring. The level of detail is a relative and not an absolute measure- a fact which is significant in later observations. A basis prerequisite for handling various levels of descriptions is the differentiation between quantitative and qualitative terms within the model. Roughly, a difference is made between adjectives to describe properties and figures to determine a required (quantitative) property.

3.2. REQUIREMENT DESCRIPTION – A THREE-STEP APPROACH

In order to be able to describe the property P named with the requirement, a procedure is needed which allows a “semantic” determination of this property in a context. The described three-step approach models this property by relating the object concerned, its property and the desired property characteristic to one another. Each of these pieces of information is referenced to the intra-logistical domain model (cf. Figure 2) in turn. Thus, the required property P is determined step by step. First the “object” O concerned is determined. Investigations into example requirements have shown that there are various rough categories for this part: the facility as a whole, certain parts and components within the facility, as well as processes and stakeholders throughout the service life of the facility. When the object O of the concerned requirement has been determined (often in the genitive form, e.g. “the velocity of the conveyor”), the “property” E(O) of object O is considered. Different aspects of the facility and its components are summarized under such a “property”; these might be either the physical features of the parts (the required width of the rolls), or a function or behaviour which the roll conveyor displays (goods are conveyed on the right-hand side of the belt). Facility parts can often be accurately named when physical properties are to be determined, e.g. “the width of the conveyor belt of the roll conveyor must be 60 cm” determines the feature “width” of the object “conveyor belt” (and the object “rolls”).

On the other hand, requirements on a function or behaviour tend to be of a more abstract nature and must be backed with suitable and meaningful concepts in the domain model. It is often necessary that the requirements processing system derives more concrete (precise) requirements from abstract requirements or presents the user with alternative interpretations, when and if possible. Such “detailed specification” is possible if either the object concerned, the desired requirement or the targeted characterization of the property is described in more detail – in compliance with the stakeholder’s intention (user interaction e.g. by means of system enquiry may be necessary).

![Figure 4: Exemplary interpretations of flexibility requirements](image-url)
When analyzing the aforementioned flexibility requirement, it allowed up to ten different interpretations using related terms which express several distinct types of “flexibility” (some of which are presented in Figure 4). This led to more detailed requirements concerning different parts of the facility: not only requirements on the roll conveyor to be used, but also some directed towards the goods to be transported and the system’s functionality in general (services). Moreover, the flexibility requirement is a good example to explain the processing and determination of properties for the data model. When handling unstructured requirements as in this case, both qualitative and quantitative types of requirements have to be supported (cf. [9] for mixed forms). Qualitative requirements can be characterized with an adjective and often the level or degree strived for (“as flexible as possible” / “highly flexible”). Contrary to this are quantitative properties which consist of one numerical value (or an interval cf. [14]), a unit and maybe a comparison operator, e.g. “> 0.5 m/s”. Quantitative properties feature the highest level of detail and can thus be used as specifications, conditions or constraints for the facility. (“Throughput must be > X units/ h!”).

4. Different Forms of Requirements Relations

Taking the levels of detail into account when processing requirements is just one aspect from the host of stakeholder requirements – albeit a major one. It is used to structure the number of requirements for further processing and to investigate their dependencies (in this case, the refinement and detailed specification of a requirement). As already described, various requirements can be referenced to one another through the level of detail. The term relation or inter-action is of particular significance in this connection. It is however not clearly used in literature dealing with requirements [2, 7, 9-11] and it can be viewed from different angles. Papers [15] and [9] explicitly include e.g. the properties of language in their investigations. Using and adapting those views, the following will be looking at the various types of relations which may occur in an intra-logistical facility. Taking these requirements relations into consideration is a major contribution to reducing time, costs and complexity in facility planning, for example by identifying requirement redundancies or refinements. Furthermore, the technical and logical connections within roll conveyors and between the conveyor and its surrounding are investigated and made explicit when the model is validated. Thus the requirement processing system is placed into a situation where it can determine conflicts when processing the requirements and in this way contribute greatly to facility design.

The term relation stands for the fact that one requirement is connected to another in a technical, logical or organizational way. The relations under investigation here may be, on the one hand, “functional dependencies” or on the other hand, “logical dependencies”. In the case of a functional relationship, the existence of one or several other elements is determined by the existence of one element, whereby the elements concerned must however be connected to one another in a structured and formalized way; e.g. the object O has exactly the same properties E(O). On the other hand, “logical dependencies” represent mostly technical connections between requirements. Their objective is to consider the actual situation in the requirement administration system already at an early stage, facts which otherwise would not be included until later on in the development. These relations originate from technical and organizational connections within the facility and its parts.

4.1. Three Different Levels of Requirement Relations

On the basis of this rough subdivision, three different types of dependencies can be identified which in each case are based on the interaction between template and domain model. The first two types of relationship occur “within a requirement”, the third describes dependencies between two or more requirements.

In the case of direct template-internal dependencies within a requirement, the value of another element is given by stating a data element (cf. functional dependency). For example, each requirement has a clear text, a creator etc., these are functional dependencies which are already imaged in the data model as such. Extended template-internal dependencies of a requirement refer on the other hand not only to the data model but are additionally based on information from the domain model of the intra-logistical facility and help to identify for example conflicts between two requirements A and B: assuming both requirements refer to object O and its property E. However, if the requirements demand characteristics u and v, and u and v are different, the requirements A and B will contradict one another.

The most important category is that of template-external dependencies between the requirements A and B imaged in the requirement template. Their relation can (if necessary, via interim steps) be due to connections of components and properties described by it, if the data model contains this logical dependency (cf. [10, 6, 7]). This type of dependency is of importance in the efficient planning of an intra-logistical facility as they help to obtain knowledge on technical connections and possible difficulties in implementation.
5.2 ANALYSIS OF RELATION CATEGORIES

The group of template-external relations underwent step-for-step refinement in further investigations. The overall number of requirements can then be better algorithmically investigated by the requirement processing system, in order to obtain a more sensible and, most importantly, a more efficient design for later implementation and applied “implementation alternatives” [11, 6]. In order to draw up a general model for template-external relations for intralogistics, a combined approach using bottom-up and top-down procedures were selected. General models on requirements relations were analyzed, compared and finally aligned to the intralogistics sector. In this case, this involved especially the approaches described in [15], [11], [9] and [8], which in parts identify various types of relations based on one another. The types of relations identified and their number were compared.

At the same time, the relations were validated, and supplementary dependencies between relations were determined using existing requirements material and by consulting experts and works of reference etc. Problems arose in limiting the types of relations. However, it must be said that this problem occurs in all comparable strategies as allocations are not always clearly made. Our own model unites the individual advantages of the strategies, for example, the support of requirement networks instead of simple requirement hierarchies. Furthermore, the model was extended to include bi-directional relations, which express direct interactions between components, functions and properties within the facility. That means, two contrary relations occur simultaneously.

5. RELATIONS IN RELATION NETWORKS

The following sections deal with the definition of those types of relations identified in the requirements analysis and, if necessary, their demarcation from one another. The starting point is consideration of the existence of numerous requirements for an intra-logistical facility which can be transferred using the template and supported by the domain model “intra-logistics” into a structured form. In order to obtain the best possible facility planning which meets as many requirements as possible, the dependencies between the requirements when implemented must be identified. All of these dependencies generate a relation network out of the unstructured requirements volume – if viewed collectively.

The basis of sub-dividing the identified types of relations lies in the existing requirements in one hierarchy i.e. upper-level and lower-level requirements are determined in respect of a prescribed relation. One of the most frequently used relations for limited problem fields is the above-mentioned relation based on the “level of detailing” (often referred to as “refined relation”) between [8, 7, 10]. However, other relations are also conceivable, as defined below. Generally, a difference can be made between vertical, horizontal and network relations when determining the interactions of requirements using the roll conveyor as an example.

5.1 VERTICAL RELATIONS

Vertical relations refer to those relations which on the one hand, are used to create a hierarchy of requirements or on the other hand, can run between the levels of a prescribed hierarchy. That means there is always a description of an upper and lower order relation between requirements so that an adjusted relation exists. A more exact distinction can be made between detailing relations and “is a part of” relations. In the case of detailing relations such as “refinement relations”, the lower-level requirement contains more information than the upper-level requirements, i.e. lower-level requirements address more technical issues. Such a “refinement” is present when either the property is expressed more exactly (“variable velocity” instead of “flexibility”; refer to Figure 4) or the value of the property is specified (“velocity of 0.7 m/s”). The latter often involves a property scale change if when specifying, the requirement changes from being qualitative to quantitative (cf. e.g. [14]). The detailing relation can however only be sensibly applied to small requirement sets or sub-sets at a time. It is therefore possible that requirements in view of the “refinement relation” cannot be compared with one another. This may be remedied by prescribing a fixed number of detail levels. However, this also means that the depth of the requirement hierarchy is limited to this number of levels. Further project work will investigate whether this means too great a restriction.

A further important relation used to generate a requirement hierarchy is the “is part of” relation, which is based on a component model as a reference. The component model – e.g. rolls, drives, belts and motors as components of a roll conveyor – shows the aggregation structure of the parts of a component according to which the requirements are hierarchically classified – in this example “roll-related”, “drive-related”, “belt-related”, etc. The requirement is classified in the hierarchy according to the object it is aligned to. Thus the requirement hierarchy is based just on the reference object and not on its properties, making it much easier to create such a hierarchy.
5.2 **Horizontal Relations and Networks**

*Horizontal relations* are the opposite of vertical relations. They do not describe upper or lower relations but existing dependencies “on one or more hierarchal levels”. Here a difference is made between qualitative and quantitative relations which take into account, the object, property and value of the requirement.

*Qualitative relations* serve to describe logical factual connections between the requirements in the template. They refer exclusively to terms from the domain model, but describe the (qualitative) impact of a requirement, rather than a mathematically correct formula. In all, four types of impact can be identified and further sub-divided (cf.[11, 8]). Either one requirement is needed to meet another (necessary relation e.g. drive-related requirements and requirements regarding energy supply), or support of two requirements (bidirectional relation, demand for higher velocity and high-level throughput), or impediment between two requirements (rival relation e.g. demand for higher level of running smoothness and higher velocity, or fourthly, contrary requirement relation (excluding relation e.g. demand for a roll width smaller than the width of the goods to be transported). All these types of impacts can be unidirectional (especially in the case of need) but also interactive (especially in the case of contradiction).

Furthermore, the influence of changes in properties on requirement B which is connected to requirement A can be qualitatively described. On the other hand, these relations of impact take into consideration four different combinations: +/- relations, in which case an improvement of a property or an increase in the value of an attribute of an initial requirement A leads to an improvement of the targeted property for B, +/- relations where an increase in one property (of A) leads to a worsening of the connected property of B; as well as the analogue -/+ and -/- relations. Figure 5 provides additional illustration of this in a diagram with corresponding annotation where requirements made on flexibility, reliability and safety and their respective connections are depicted. For example, in the case of a roll conveyor a higher number of goods can be moved in the same time when the operating speed of the conveyor is increased. There is therefore a bidirectional relation of type +/- between the qualitative requirements on the possible throughput and the transportation speed of the conveyor belt.

Quantitative relations are contrary to qualitative relations and occur where property characteristics are given as a number (often including a unit). If a mathematical dependency between two quantitative existing requirements can be defined using standard methods of algebra, then there exists a quantitative relation.

“Network relations” exist as an additional and the most general group of requirement relations, which serve to model complex expert knowledge. Requirements can be shown as a network in which, on the one hand, several relations can be viewed simultaneously and, on the other hand, the impact of one requirement in a branch of the hierarchy on requirements in other branches can be investigated. Vertical relations in a network are still acyclic and induce the usual tree-like branching between one requirement on one hand and one or more requirements on the other hand (cf. hierarchy). Horizontal relations of quantitative and qualitative nature result however in a comprehensive impact graph for the effect of individual requirements and, if required, for the effects and interactions between the required properties among themselves.

Purpose-means relations as an additional aspect regarding the volume of requirements can serve as another example. The initial requirement A represents a means of fulfilling a certain purpose, the fulfillment of target requirement B whereby whole chains of purpose-means relations between requirements over several hierarchy levels are conceivable. Furthermore, it holds that a “means requirement” can serve several “purposes” and that a “purpose” can be allocated to several “means”. This extends the direct reference of qualitative impact relations with the possibility to determine “cardinality” for the effects.
The resulting complex relation networks which can already be identified in Figure 5 serve to illustrate expert knowledge and physical-technical and socio-economical dependencies. As the examples show, the structured recording of these networks allows requirements and their impacts to be taken into optimal consideration when planning the facility.

6. CONCLUSION AND FUTURE WORK

The given structuring of requirement relations has used those concepts for a requirement model and the domain ontology for intra-logistics drafted so far in the project in order to be able to identify at as early a stage as possible requirements dependencies. The first sub-targets of the project have been reached i.e. requirements and their relations have been transferred into a model and knowledge about the specialist domain can also be additionally imaged. This is advancing an efficient planning system for intra-logistical facilities. Further steps of the project will be concerned with the development and testing of additional methods for interactive term integration in the domain model. With regard to the desired implementation of requirements in a rough concept, implementation alternatives need to be identified and integrated into the planning system via computer assistance. Furthermore, dependencies between individual “parts” of the intra-logistical facility must be investigated and formalized.

The authors wish to thank the DFG (Deutsche Forschungsgemeinschaft) for its support of this project.

REFERENCES

Characteristics of Successful Kaizen Events: An Empirical Study

Jennifer A. Farris¹, Eileen M. Van Aken², Toni Doolen³, June Worley³, and Wen-Hsing Liu¹

¹Department of Industrial Engineering
Texas Tech University
Lubbock, TX, 79409, USA

²Grado Department of Industrial & Systems Engineering
Virginia Tech
Blacksburg, VA, 24061, USA

³Department of Mechanical, Industrial, & Manufacturing Engineering
Oregon State University
Corvallis, OR, 97331, USA

ABSTRACT

Kaizen events are short-term improvement projects focused on producing accelerated change in a targeted work area or process through the implementation of lean manufacturing tools and techniques. Kaizen events share several characteristics with more "traditional" quality improvement activities, such as the use of root cause analysis and work measurement tools. However, they differ significantly in other ways, including their accelerated timeframe, their high degree of implementation authority, and their focus on cycles of implementation and "hands on" experimentation during the event. Kaizen events are widely used in many different industries, and anecdotal evidence indicates that they have resulted in significant improvements within some organizations. However, Kaizen events have received limited research attention to date, and it is not fully known what factors contribute most to event effectiveness, either short-term or longer-term. This paper presents results from a field study of 51 Kaizen events in six US manufacturing organizations. An initial model of event effectiveness was developed based on literature in related areas (teams, projects, organizational change). Multiple methods, including regression modelling, observation, and qualitative data analysis, were used to understand event operation and to identify the determinants of both human resource and business performance outcomes. This paper presents results from two analyses on the same data set using different methods, a quantitative study using regression modelling and a qualitative study of responses to open-ended survey items, in order to draw insight about what factors appear to be most strongly related to event effectiveness. In addition, we find that the results of the qualitative and quantitative analyses do not fully coincide, suggesting that participant mental models of the factors that contribute to success may be incomplete. The paper concludes by identifying opportunities for future work.

1. INTRODUCTION

"Kaizen events" are focused and structured continuous improvement projects, which use dedicated cross-functional teams to achieve rapid improvements in a targeted work area in an accelerated timeframe (usually one week or shorter) [1]. Although they appear to have originated with Toyota in the 1970s [2], they do not appear in the manufacturing literature until the mid 1990s [3], at which time they began to spread rapidly in manufacturing organizations worldwide, often in association with lean manufacturing programs. Reported results from Kaizen events include such “breakthrough” performance improvements as a single-year productivity increase of nearly nine hundred percent, a work-in-process (WIP) reduction of 89%, and a cycle time reduction of 79% in one work area at AlliedSignal Inc.'s Phoenix jet engine plant [2]. In addition, one practitioner resource has suggested that a “typical” Kaizen event in a manufacturing process results in a 20-60% increase in productivity, a 40-80% decrease in WIP, and a 30-70% reduction in cycle time [4]. However, these, and most other, published results are anecdotal. There is currently a lack of theoretical understanding of what factors determine event effectiveness. Meanwhile, Kaizen events are sufficiently different than previous continuous improvement mechanisms, e.g., quality circles and continuous process improvement teams, to warrant a separate investigation. For instance, in Kaizen events, the teams often have authority to implement changes during the event, creating the potential for rapid, immediate
improvement in the targeted work areas, as well as increased employee buy-in to event effectiveness [5]. On the other hand, the accelerated timeframe of the events means that employees may have less time to identify, plan, and test improvement strategies. Of even greater concern, perhaps, is whether the improvements initially generated can be sustained in the longer-term.

This paper contributes to the existing body of knowledge of Kaizen event effectiveness by comparing the results of two analyses focused on identifying the determinants of event effectiveness. Both analyses were conducted on the same data set of 51 Kaizen events from six US manufacturing organizations. However, different input data were used. The first analysis used regression modelling to identify statistical relationships between Kaizen event input, process, and outcome factors. The second analysis was a qualitative study of responses to open-ended survey items. We find that the results of the qualitative and quantitative analyses do not fully coincide, suggesting that participant mental models of the factors that contribute to success may be incomplete. The remainder of this paper is organized as follows: Section 2 presents additional background on the study motivation and design. Section 3 presents results from the first study, while Section 4 presents results from the second study. Finally, Section 5 compares the results of the two studies, and Section 6 identifies opportunities for future research.

2. BACKGROUND

2.1. RESEARCH MODEL

Previous research on Kaizen events has been limited and has consisted primarily of small sample case studies. Two notable exceptions are Bateman [6] who studied the determinants of process improvement sustainability in 40 process improvement projects from 21 organizations and Miller [7] who studied the relationship between Kaizen program participation and employee motivation outcomes in 166 individuals from four organizations. The existing research on Kaizen events, the practitioner literature on Kaizen events, and related organizational theory, e.g., on projects and teams, were all used to develop the initial model for this research (see Figure 1). Due to space constraints, detailed justification of the model and the variable interrelationships is omitted here but can be found in [8]. In essence, the Kaizen event practitioner literature and organizational theory suggest first, that two different types of outcomes are of interest in studying Kaizen events: technical system outcomes related to goal achievement, and social system outcomes related to human resource development. Second, the Kaizen event practitioner literature and previous research suggest that the input factors and process factors identified in Figure 1 are among the most likely key determinants of the Kaizen event outcomes. Although support for some of the relationships is stronger than others, there is reason to believe that any of the input factors may impact the outcomes either directly or through the Kaizen event process factors, and that any of the Kaizen event process factors may be linked to any of the outcomes. In addition, although not tested in this research, the literature suggests that organizational and work area context factors may influence Kaizen event design factors. The research hypotheses are summarized in Table 1, while Table 2 provides definitions of the factors.

![Initial research model](image)
**Characteristics of Successful Kaizen Events: An Empirical Study**

Table 1: Research hypotheses

<table>
<thead>
<tr>
<th>Hypothesis Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis H1</td>
<td>Input variables have a direct relationship to outcomes.</td>
</tr>
<tr>
<td>Hypothesis H2</td>
<td>Kaizen event process variables have a direct relationship to outcomes.</td>
</tr>
<tr>
<td>Hypothesis H3</td>
<td>Kaizen event process variables partially mediate the effect of input variables on outcomes.</td>
</tr>
</tbody>
</table>

Table 2: Factor definitions

<table>
<thead>
<tr>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal Clarity</td>
<td>Team perceptions of the clarity of their improvement goals</td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>Team perceptions of the difficulty of their improvement goals</td>
</tr>
<tr>
<td>Team Autonomy</td>
<td>Team perceptions of the amount of control over event activities given to the team</td>
</tr>
<tr>
<td>Team Kaizen Experience</td>
<td>Average number of Kaizen events team members have participated in</td>
</tr>
<tr>
<td>Team Leader Experience</td>
<td>Number of Kaizen events the team leader has led or co-led</td>
</tr>
<tr>
<td>Team Functional Heterogeneity</td>
<td>Diversity of functional expertise within the Kaizen event team</td>
</tr>
<tr>
<td>Management Support</td>
<td>Team perceptions of the adequacy of resources dedicated to the event</td>
</tr>
<tr>
<td>Event Planning Process</td>
<td>Total person-hours invested in planning</td>
</tr>
<tr>
<td>Work Area Routineness</td>
<td>General complexity of the target system, based on the stability of the product mix and degree of routineness of product flow</td>
</tr>
<tr>
<td>Action Orientation</td>
<td>Team perceptions of the extent to which their team focused on implementation versus analysis</td>
</tr>
<tr>
<td>Affective Commitment to Change</td>
<td>Team perceptions of the need for the Kaizen event</td>
</tr>
<tr>
<td>Internal Processes</td>
<td>Team perceptions of the internal harmony and coordination of their team.</td>
</tr>
<tr>
<td>Tool Quality</td>
<td>Facilitator perceptions of the quality of the team’s use of problem-solving tools</td>
</tr>
<tr>
<td>Tool Appropriateness</td>
<td>Facilitator perceptions of the appropriateness of the problem-solving tools used by the team during the event</td>
</tr>
<tr>
<td>Understanding of Continuous Improvement</td>
<td>Team perceptions of their incremental gains in knowledge of continuous improvement resulting from a specific Kaizen event</td>
</tr>
<tr>
<td>Skills</td>
<td>Team perceptions of their incremental gains in problem-solving skills resulting from a specific Kaizen event</td>
</tr>
<tr>
<td>Attitude</td>
<td>Team perceptions of the degree to which members gained affect for events from a specific Kaizen event</td>
</tr>
<tr>
<td>Goal Achievement</td>
<td>Aggregate percentage of major improvement goals met</td>
</tr>
<tr>
<td>Impact on Area</td>
<td>Team perceptions of the impact of the event on the target area</td>
</tr>
<tr>
<td>Overall Perceived Success</td>
<td>Facilitator perceptions of the overall success of the Kaizen event.</td>
</tr>
</tbody>
</table>

**2.2. RESEARCH METHODOLOGY**

Following the creation of the research model, a field study was used to test the hypothesized interrelationships at the team level of analysis. Data were collected from 55 Kaizen events in six US manufacturing organizations; however, four events were ultimately removed from the analysis due to incomplete data, leaving the final sample size of 51. Within each organization, events were randomly selected using either a census or a systematic sampling approach. However, organizations were non-randomly selected due to the need for repeat access to organizations. The selection criteria were organization type (manufacturing), experience in conducting Kaizen events (at least one year), systematic use of Kaizen events (part of organizational strategy) and frequent use of Kaizen events (at least one per month). The criteria were intended to reduce variability across organizations due to learning curve effects and differences in organization strategy, as well as to allow a sufficient sample size within each organization. The selected organizations come from a variety of industries (see Table 3). Data were collected over a nine month time period from October 2005 through June 2006.

Given the objectives of the study, measures were devised for each of the 20 variables in Figure 1 and Table 2. A mixture of objective and perceptual measures was used. The data collection process for each event was as follows. First, on the first day of the Kaizen event, immediately following the kickoff meeting, participants were asked to
complete a short questionnaire describing their perceptions of their team goals (19 items). Second, on the final day of the event, immediately following the report out meeting where the team presented results to management, participants were asked to complete a second questionnaire (39 items) describing their perceptions of other input and process variables, i.e., management support, team autonomy, internal processes, action orientation, as well as four outcome variables, i.e., understanding of continuous improvement, skills, attitude and impact on area. This survey also included two open-ended questions asking participants to identify what they perceived to be the biggest obstacles and biggest contributors to event success. These open-ended questions were intended to allow the identification of possible additional influential factors, as well as to allow comparison of participant mental models of success to regression analysis results. Finally, the facilitator of the event was asked to complete a questionnaire (15 items) to capture data on additional variables, some objective (e.g., team composition and team success versus goals) and some perceptual (e.g., facilitator rating of the overall success of the team). In addition, the facilitator was asked the same two questions regarding biggest obstacles and biggest contributors to event success.

All questionnaires were constructed according to standard principles of questionnaire design [9], and constructs were based on previously validated instruments whenever possible. Questionnaires were pilot tested before the research and exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and Cronbach’s alpha were used to validate the measures. In general, the EFA and CFA results supported the construction of the measures, although in some cases the removal of one or more items from a given scale was supported (see [8] for more detail). The most significant change was the combination of the Understanding of Continuous Improvement and Skills scales into a single scale, named Kaizen Capabilities, which can be defined as “Team member perceptions of the incremental gains in their problem-solving knowledge, skills, and abilities resulting from a specific Kaizen event.” That is, although most learning frameworks suggest that knowledge and skills represent different dimensions, this research suggests that they are highly related empirically. All scales had a Cronbach’s alpha of 0.70 or greater except for the Action Orientation scale, which had an alpha value of 0.64, which falls into the acceptable range for newly-developed scales [10]. Following the validation of the measures, the final step was to determine whether teams demonstrated sufficient within-group agreement on measures to make team-level aggregation meaningful. ICC(1) and rwg were used to test the consistency of responses within versus across teams and aggregation was supported in all cases (see [8] for additional detail).

Table 3: Study organizations

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Secondary wood products manufacturer</td>
<td>Electronic motor manufacturer</td>
<td>Secondary wood products manufacturer</td>
<td>Manufacturer of large transportation equipment</td>
<td>Specialty equipment manufacturer</td>
<td>Steel component manufacturer</td>
</tr>
<tr>
<td><strong>SIC code</strong></td>
<td>2434</td>
<td>3621</td>
<td>2434</td>
<td>3731</td>
<td>3843</td>
<td>3443</td>
</tr>
<tr>
<td><strong>Year founded</strong></td>
<td>1946</td>
<td>1985</td>
<td>1946</td>
<td>1939</td>
<td>1964</td>
<td>1913</td>
</tr>
<tr>
<td><strong>Number of employees</strong></td>
<td>560</td>
<td>700</td>
<td>500</td>
<td>18,000</td>
<td>950</td>
<td>3500</td>
</tr>
<tr>
<td><strong>Event rate during research</strong></td>
<td>2–3 per month</td>
<td>1 per month</td>
<td>2 per month</td>
<td>5-6 per month</td>
<td>6-8 per month</td>
<td>1 per month</td>
</tr>
<tr>
<td><strong>Number of events sampled (retained)</strong></td>
<td>15(15)</td>
<td>8(8)</td>
<td>11 (7)</td>
<td>4(4)</td>
<td>12(11)</td>
<td>6(6)</td>
</tr>
</tbody>
</table>

3. STUDY ONE

Following a check of variance inflation factors and distributional properties of the variables, regression analysis was used to identify the direct and indirect predictors of each outcome. Given the lack of previous research to suggest which factors were most important, an exploratory regression was used to develop a model for each outcome. A backward selection procedure was used, removing the variable with the largest p-value until all remaining variables in the model were significant at the $\alpha = 0.10/k$ level, where k is the number of parameters in the
model. Then, using input from auxiliary search methods, e.g., automated SAS searches using $C_p$, $R^2$ and adjusted $R^2$, additional models were considered where some variables had $p$-values between 0.10/k and 0.05. Due to the nested structure of the data, it could not be assumed that the responses for teams within a given organization were uncorrelated [11]. Therefore, generalized estimating equations (GEE) [12], executed in SAS 9.1.3 PROC GENMOD, were used to account for correlation between teams within the same organization, which may bias the estimates of parameter standard errors and associated F-tests [13,14]. However, ordinary least squares approximation was used to further confirm the robustness of modelling results, e.g., through the auxiliary search methods. Table 4 summarizes the regression modelling results for the five outcome variables. It should be noted here than in Model 4, logistic regression was used, whereas in the other four models, Gaussian regression was used.

Following identification of the direct predictors of outcomes, mediation analysis was used to test the hypothesized indirect relationships. Only process variables found to be significant in the previous regression (see Table 4) were tested as potential mediators. A modified version of Judd and Kenny’s [15] classic approach was used as follows. In Step 1, the mediating process variable (Z) was separately regressed on each of the nine input variables (X) and the resulting coefficient was tested for significance. If a significant relationship was demonstrated in step one, in Step 2, the outcome variable (Y) was regressed on both the input variable (X) and the mediating process variable (Z), and the resulting regression coefficients were tested for significance. A significant relationship between the mediating process variable (Z) and the outcome (Y) suggests a mediation relationship. The relationship between the outcome variable (Y) and the input variable (X) can be either significant (partial mediation) or non-significant (full mediation). Since each regression coefficient was tested separately, a Bonferroni correction ($\alpha = 0.05/3 = 0.0167$) was used to account for the increased chance of type I error in these first two steps [16]. After the two preceding steps were accomplished for all nine input variables, in Step 3, the mediating process variable (Z) was simultaneously regressed on all the input variables (X) significant in step one. This step was performed to confirm whether each input variable (X) was a significant unique predictor of the mediator (Z), after controlling for the other input variables. Table 5 displays the significant mediation relationships discovered.

4. STUDY TWO

In the second of the two studies, qualitative analysis methods [17] were used to analyze participant responses to the open-ended questions on the biggest obstacles and contributors to team success. A total of 51 facilitators and 301 team members responded to the open ended questions, resulting in 289 individual-level responses for the obstacles question and 379 individual-level responses for the contributors question. Two of the authors coded the data using a code list developed by one of the two coders. The initial coding reliability was 67% for the obstacles question and 71% for the contributors question, while the code-recode reliability was 89% and 88%, respectively. The final reliability was 100%, as, after discussion, the two coders were able to come to agreement on all responses. Following initial coding, the lower-level codes were then grouped into 17 clusters representing five overall themes. Table 6 presents the cluster list, while Table 7 displays Pareto analysis results by response frequency for obstacles and contributors.

The lists of most frequent obstacles and contributors are similar, but not identical, suggesting that participant mental models of the contributors to success are slightly different than participant mental models of the obstacles to success. First, it is noted that process characteristics and process contributor support appear only in the obstacles lists, while internal processes appears only in the contributors list. It is also noted that the top three response categories (roughly 60% of responses in each case) have a very different nature. The top three response categories for obstacles are technical factors, two of which are clearly external to the team, i.e., process characteristics and resource support. Meanwhile, the top three response categories for contributors are social factors related to the team and team dynamics. This may be due to the fundamental attribution error, which is the tendency for people to attribute positive outcomes to factors they can control, while attributing negative outcomes to uncontrollable or external factors [18].

5. COMPARISON OF STUDY RESULTS

Comparison of the results of the two studies generates several interesting observations. The first is the relative congruence of the qualitative analysis results with the regression analysis results. All of the factors in Table 7 appear in some form as significant factors in the regression models. This may suggest that participant mental models possess some degree of accuracy. However, this correspondence may also be due, at least in part, to same source bias, as many of the variables used in the regression models were perceptual and measured on participants.
Second, if relative accuracy of participant mental models can be assumed, comparison of the regression results with the qualitative analysis results suggests that participants define success in a broader sense than simple achievement of technical outcomes. Several of the most frequent responses in the qualitative analysis appear only in the human resource outcome models, i.e., Internal Processes, Affective Commitment to Change, and Team Composition. This further supports the need to consider both technical and human resource outcomes in research and evaluation of Kaizen events. Third, although participant mental models appear to be fairly accurate, this research suggests that participants underestimate the impact of Goal Clarity, Goal Difficulty, and Team Autonomy, which were significant factors in the regression results, but do not appear among the top factors in the qualitative analysis results. All three variables were involved in indirect relationships to outcomes, although Goal Difficulty and Team Autonomy were also directly related to outcomes, perhaps suggesting why participants failed to identify their influence. Fourth, conversely, it appears that participants may overestimate the impact of Affective Commitment to Change, which appeared only in the Kaizen Capabilities model and was only marginally significant.

Table 4: Regression modeling results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1 y = Attitude</th>
<th>Model 2 y = Kaizen Capabilities</th>
<th>Model 3 y = Impact on Area</th>
<th>Model 4 y = Goal Achievement</th>
<th>Model 5 y = Overall Perceived Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>β = 0.467 p = .447</td>
<td>β = 0.219 p = .687</td>
<td>β = 0.953 p = .126</td>
<td>β = 17.383 p = .009</td>
<td>β = 1.997 p = .047</td>
</tr>
<tr>
<td>Team Functional Heterogeneity</td>
<td>-0.547 .012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Support</td>
<td>0.250 .013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Processes</td>
<td>0.694 &lt;.001</td>
<td>0.465 &lt;.001</td>
<td>0.262 .046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>0.119 .032</td>
<td></td>
<td></td>
<td>-1.979 .043</td>
<td></td>
</tr>
<tr>
<td>Team Autonomy</td>
<td>0.234 .014</td>
<td>0.342 .022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team Kaizen Experience</td>
<td>-0.398 &lt;.001</td>
<td></td>
<td></td>
<td>-4.254 .003</td>
<td></td>
</tr>
<tr>
<td>Team Leader Experience</td>
<td>-0.195 .020</td>
<td></td>
<td></td>
<td>-4.262 .001</td>
<td></td>
</tr>
<tr>
<td>Work Area Routineness</td>
<td>0.094 .002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective Commitment to Change</td>
<td>0.222 .049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action Orientation</td>
<td>0.243 &lt;.001</td>
<td>-1.901 .028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event Planning Process</td>
<td>4.554 .002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tool Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.621 .004</td>
</tr>
<tr>
<td>R²</td>
<td>0.605</td>
<td>0.706</td>
<td>0.466</td>
<td>N/a</td>
<td>0.123</td>
</tr>
<tr>
<td>R²a</td>
<td>0.580</td>
<td>0.658</td>
<td>0.432</td>
<td>N/a</td>
<td>0.105</td>
</tr>
<tr>
<td>p</td>
<td>-0.019</td>
<td>-0.071</td>
<td>-0.071</td>
<td>-0.071</td>
<td>-0.057</td>
</tr>
</tbody>
</table>

Table 5: Significant mediation relationships

<table>
<thead>
<tr>
<th>Input variable</th>
<th>Mediating process variable</th>
<th>Outcome variable</th>
<th>Estimated effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal Clarity</td>
<td>Internal Processes</td>
<td>Attitude</td>
<td>0.414</td>
</tr>
<tr>
<td>Goal Clarity</td>
<td>Internal Processes</td>
<td>Kaizen Capabilities</td>
<td>0.348</td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>Action Orientation</td>
<td>Impact on Area</td>
<td>-0.171</td>
</tr>
<tr>
<td>Team Autonomy</td>
<td>Action Orientation</td>
<td>Impact on Area</td>
<td>0.182</td>
</tr>
<tr>
<td>Work Area Routineness</td>
<td>Action Orientation</td>
<td>Impact on Area</td>
<td>0.106</td>
</tr>
<tr>
<td>Management Support</td>
<td>Tool Quality</td>
<td>Overall Perceived Success</td>
<td>0.316</td>
</tr>
</tbody>
</table>
### Table 6: Clusters for qualitative analysis

<table>
<thead>
<tr>
<th>Themes</th>
<th>Clusters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizational Context</td>
<td>Management Commitment</td>
<td>Willingness of senior executives and lower-level managers to accept change or</td>
</tr>
<tr>
<td></td>
<td>Supporting Departments</td>
<td>provide help</td>
</tr>
<tr>
<td></td>
<td>Resource Support</td>
<td>Willingness of organizational department other than the target area to accept</td>
</tr>
<tr>
<td></td>
<td>Environment</td>
<td>change or provide help</td>
</tr>
<tr>
<td>Kaizen Event Design</td>
<td>Goal Clarity</td>
<td>Extent to which event goals are clear to team</td>
</tr>
<tr>
<td></td>
<td>Goal Difficulty</td>
<td>Extent to which the event was challenging due to the scope of the goals set</td>
</tr>
<tr>
<td></td>
<td>Team Composition</td>
<td>Team compositional characteristics which affected the success of the event</td>
</tr>
<tr>
<td></td>
<td>Team Autonomy</td>
<td>Extent to which the team was given control over its work during the event</td>
</tr>
<tr>
<td>Work Area Context</td>
<td>Process Characteristics</td>
<td>Technical characteristics of the target work area which affected team progress</td>
</tr>
<tr>
<td></td>
<td>Process Contributor Support</td>
<td>Willingness of employees in the target work area to accept change or provide</td>
</tr>
<tr>
<td>Kaizen Event Process</td>
<td>Solution Process</td>
<td>Characteristics of the solution development and implementation process which</td>
</tr>
<tr>
<td></td>
<td>Internal Processes</td>
<td>Extent to which team member interactions were harmonious</td>
</tr>
<tr>
<td></td>
<td>Affective Commitment to Change</td>
<td>Team member commitment to the extent to which they accept change and provide</td>
</tr>
<tr>
<td>Other</td>
<td>Sustainability</td>
<td>Characteristics which affect sustaining of event results</td>
</tr>
<tr>
<td></td>
<td>Everything</td>
<td>Respondent states everything or everyone was a contributor or obstacle</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>Respondent states there were no contributors or obstacles</td>
</tr>
<tr>
<td></td>
<td>Survey Error</td>
<td>Response is non-existent or unusable</td>
</tr>
<tr>
<td></td>
<td>Non-Contributor</td>
<td>Response does not directly describe a contributor</td>
</tr>
</tbody>
</table>

### Table 7: Qualitative analysis results

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>Percentage of Responses</th>
<th>Cumulative Percentage</th>
<th>Contributor</th>
<th>Percentage of Responses</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution process</td>
<td>22.8</td>
<td>22.8</td>
<td>Affective commitment to change</td>
<td>21.4</td>
<td>21.4</td>
</tr>
<tr>
<td>Process characteristics</td>
<td>19.7</td>
<td>42.6</td>
<td>Team composition</td>
<td>20.1</td>
<td>41.4</td>
</tr>
<tr>
<td>Resource support</td>
<td>18.3</td>
<td>60.9</td>
<td>Internal processes</td>
<td>16.6</td>
<td>58.0</td>
</tr>
<tr>
<td>Team composition</td>
<td>7.3</td>
<td>68.2</td>
<td>Solution process</td>
<td>14.2</td>
<td>72.3</td>
</tr>
<tr>
<td>Supporting departments</td>
<td>5.9</td>
<td>74.0</td>
<td>Resource support</td>
<td>10.6</td>
<td>82.8</td>
</tr>
<tr>
<td>Affective commitment to</td>
<td>4.8</td>
<td>78.9</td>
<td>Supporting departments</td>
<td>5.0</td>
<td>87.9</td>
</tr>
<tr>
<td>Process contributor</td>
<td>4.5</td>
<td>83.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Future Research

The two studies described in this paper revealed several factors which appear to be related to the technical performance and human resource outcomes of Kaizen events. Furthermore, participant mental models of event success, as demonstrated through the qualitative analysis, were found to correspond highly, but not completely, with regression modelling results. Although this study provides initial insight into the factors determining event success, additional research with a larger sample of events is needed to confirm study findings. In addition, given acknowledged difficulties with sustainability of results, additional research on sustainability is also needed.

Acknowledgements

This research was supported by a National Science Foundation grant (DMI 0451512) awarded to the second and third authors. The authors also gratefully acknowledge the support of all facilitators and team members who participated in this research.

References

Developing a Business Improvement Strategy based on self-assessments using the EFQM Model

Luis Rocha-Lona1, José Arturo Garza-Reyes2, Horacio Soriano-Meier3, Tim Peacock3, Paul Forrester4

1 Engineering Systems Department
The National Polytechnic Institute, Mexico 03100

2 School of Technology
The University of Derby
Derby, DE22 3AW, UK

3 Northampton Business School
The University of Northampton
Northampton, NN2 7AL, UK

4 Birmingham Business School
The University of Birmingham
Birmingham, B15 2TT, UK

ABSTRACT

The use of the EFQM model and other business excellence models (BEM’s) has been recognized as a way of improving business processes among organizations. The information gathered through self-assessments using these models show the performance of organizations against several model criteria. There is evidence that organizations find it difficult to integrate the information from self-assessments to current business improvement programs or projects. Additionally, it has been claimed that the final reports derived from the deployment of self-assessments are rarely followed up. Thus, despite good efforts and some knowledge gained through the use of BEM’s there is a lack of methods, models, and techniques that effectively integrate the self-assessments with current business improvement strategies.

In this context, the effective deployments of BEM’s and follow up activities and plans can contribute to support and develop business improvement strategies. Under this perspective, this paper proposes an appropriate method than can help to integrate self-assessments outcomes based on the EFQM model into a business improvement strategy. Based on emerging issues derived from an empirical study with twelve European Organizations that have used the EFQM model for more than five years, the paper identifies some of the best practices to integrate self-assessment outcomes with business improvement strategies. It then proposes a framework that can help to accomplish this integration and to mitigate the problems mentioned earlier. The conclusion of this work emphasizes the necessity to standardize this process, and to integrate it with current knowledge management projects to store and retrieve the information for future business improvements projects.

1. INTRODUCTION

Business Excellence Models (BEM’s) are quality–management frameworks based on organizational performance criteria that originated as a result of the evolution of TQM principles. The BEM’s have played a significant role in the attempt to improve business among organizations, and these efforts are well documented with Quality Foundations that administer BEMs’ across regions and countries2. The models have experienced an important evolution since their introduction in the late 80’s; not only in their business-criteria but in the way they are deployed and used. In this context, organizations have learned from the use and practice of these frameworks to apply the BEM’s for several purposes. The purposes vary according to the priorities of the organization, and some of these purposes identified are award participation, self-assessment, business process improvement, measurement systems, and strategic planning [1]. This paper will consider the self-assessment, the business improvement and strategic planning approaches which are closely involved to develop business improvement strategies.

1 Corresponding author: Tel.: (+52) 5557296000 EXT. 61642; Fax: (+52) 1642-342482; E-mail: luis.rochalr@yahoo.co.uk
2 See for instance the websites of the European Foundation for Quality Management, the National Institute of Science and Technology, The Japanese Institute of Scientist and Engineers, the Canadian Quality assurance Institute, among others.
The BEMs have managed different categories to facilitate organizations assessing their own business in terms of specific criteria in their industry. Initially those categories were better suited for large public and private organizations. However, the necessity to include and expand the BEM’s to most industrial sectors encouraged Quality Foundations to develop the frameworks to other kind of organizations. Thus, the introduction of new categories to the frameworks such as health care, non-profit, education, medium and small organizations, helped enormously to increase the use of BEM’s. Figure 1 shows for instance that applications for the Malcolm Baldrige National Quality Award (MBNQA) have increased in the last years after having a setback in 1997. This increase may directly respond to the introduction of the new categories in the late 90’s. Thus, it is reasonable to think that the use of BEM’s may continue to grow as the Quality Foundations continue innovating the frameworks by industrial sectors or specific product and services.

This paper will consider the EFQM model as a framework for analysis and the discussions and conclusions of this work may be extrapolated to other BEM’s. The EFQM model is currently managed and administered by the European Foundation for Quality Management (EFQM)\(^3\). This model has become popular across the public and private sectors and it is estimated that 76 countries operate a national excellence model to promote quality improvement [4]. It is also currently estimated that approximately 30 000 European organizations employ this framework as a way of improving business and operations [4]. The EFQM model is a non-prescriptive quality framework based on 9 criteria: five enablers and 9 results [5]. The enablers are leadership, people, policy & strategy, partnerships & resources, and processes. On the other hand “results” include people results, customer results, society results, and key performance results.

As stated by Tito Conti\(^4\)[6], the EFQM model was thought and created with a systemic approach that attempt to reflect the business activity on the left hand and the “business results” on the right. Some studies have tried to prove somehow these relationships [7] [8], and they provide a good insight to understand enablers and results. However,

---

\(^3\) See the website at http:\/www.efqm.org

\(^4\) Tito Conti is one of the developers of the EFQM Model. He has written books on self-assessment and published several articles in the field. One of his most recent articles describes how the EFQM Model was born, along with some insights of it evolution. See [6]
Developing a Business Improvement Strategy based on self-assessments using the EFQM Model

there is still a need that clearly shows the relationships of enablers and results, and more importantly, to understand
with solid modeling techniques the cause and effects of enablers and their impact of business results. Under this
perspective, the EFQM model intends to provide organizations with a framework to deploy empirically the model
and assess the business. The “RADAR-logic” (Results, Approach, Deployment, Assessment, and Review) is the tool
designed to provide a practical sense when assessing business operations and functions. The RADAR helps to
establish what “results” organizations need to systematically accomplish its objectives. Thus, the RADAR suggests
the way in which organizations should plan, develop and deploy improvement methods and tools to reach the
desired objectives. In summary, the model presented in Figure 2, the RADAR logic, the fundamental concepts of
excellence, along with the model-criteria and sub criteria, constitute an overall view of the EFQM model.

2. THE EFQM MODEL AND ITS STRATEGIC ROLE FOR BUSINESS IMPROVEMENT

The use of the EFQM model to identify improvement areas is one of the main benefits of self-assessment [9],
[10], [11]. Lascelles and Peacock [12] were perhaps the first to visualize the application of the EFQM model to
strategically improve “enablers”, arguing that the “results” section of the EFQM model should set the agenda for
continuous improvement. Although it was clear the potential application of the EFQM model in this area, it was not
clear the way to accomplish it. In this way, there have been few attempts to provide approaches and methodologies
or suggest concrete improvement programs that link self-assessment outputs directly to the continuous improvement
process. The information gathered through self-assessments in the form of reports is passed on to top management
for its analysis and further use [13], but with no way to know further actions. The process ends with these reports,
and consequently, it is the ability of top management to decide what areas are priorities and how to improve those
areas through specific improvement programs. Thus, the success of this process may be limited to the correct
interpretation of top management, and the available guidance to effectively use the self-assessment outcomes.

![Business Improvement Strategy Diagram]

Figure 3: Business performance improvement feedback. Source: [12], p.110.

From figure above it is reasonable to think that organizations can integrate feedback derived from self-
assessment activities to current business plans. After conducting self-assessment organizations should be in position
to support planning activities at operational and strategic levels. The emerging issue is that there is little empirical
evidence that shows how organizations that have used the EFQM model manage to follow up self-assessment
outcomes. The only available evidence is when managers tell how they deployed the EFQM model and the real
benefit for identifying areas of improvement at operational or strategic level. For instance, a manager for a large
company in Europe said: ‘Along the way, the model helped us to identify what we were doing well, but also where
we could improve. It looked for continuous improvement which was across always…” Another manager commented “… [We made a full assessment of the organization and there were a lot of things that we wanted to improve]. [We created a list of 20 points. Then we prioritize those points, we made a score for each of them, so then we said: very important things are the activities in red color, then a yellow label, the things that are all right could be improved (green color)...” Finally, another one commented “As a result of the use of the model, we realized that the most important area for improvement was our people. Then, after analyzing the problem, we decided to deploy two concepts: communication and empowerment…”

These comments give interesting insights on how managers use strategically the EFQM model to identify areas of improvement. However, as mentioned earlier there still a gap to integrate this outcomes to current business strategy, which should address a more standardized approach to accomplish this objective. Thus, the fact those organizations are able to identify areas of improvement neither warranty improvements nor their effective ways to accomplish them.

3. LINKING SELF-ASSESSMENT PROCESS AND THE BUSINESS STRATEGY

Self-assessment provides organizations with a “picture” of their business processes on a regular basis, and helps to identify areas for improvements. However, self-assessment requires discipline and objectivity to conduct the process and to interpret the results of these activities objectively. Consequently, some organizations use external services to assure that the outcomes of this process accurately reflect the state of the business. In this way, self-assessment is widely accepted as a systematic and regular view of the organizations’ activities [9]. The self-assessment process implies that organizations employ the EFQM criteria to periodically monitor their business activity, particularly the performance of core processes. The use of self-assessment has expanded in parallel with the use of BEMs, and there is evidence that many organizations use this process in Europe [14]. However, like any other approach, there are implementation problems and challenges to measure the results of these activities. Consequently, this has originated criticism as well as a tendency to question the value of this process and to justify the resources assigned to these projects by organizations.

Like TQM concepts, self-assessment lacks a structure or methodology that can be uniformly implemented, and relies on suggestions and some tools provided by Quality Foundations and some researchers [2], [15] and [16]. Consequently, many consulting firms and professionals offer their services claiming to have “the best recipe” for implementing this process. This may help organizations at the beginning; however, more desirable results will come to organizations that take control of their own activities in the long-term [11]. Many problems also have been identified; [17] offers a good classification of these problems, which are mainly related to leadership issues and the lack of structure of the “methodology” for identifying “where to start”. Apart from leadership and commitment, these problems seem to be closely related to the way of conducting self-assessment (methodology) and the interpretation and use of the outcomes of this process. Hence, it seems that there is little evidence of standard
Developing a Business Improvement Strategy based on self-assessments using the EFQM Model

methodologies for conducting self-assessment and for integrating these results in further stages of development for the benefit of the organizations [9].

Figure 5 shows the model that concentrates rich data obtained from organizations that have used the EFQM model for several years. According with the set of organization, they usually construct instruments\textsuperscript{5} to collect information of the self-assessment process. Then, from this collection process, they analyze the information, which results in an internal/external diagnostic of the organization that can be categorized by area or function. Then, this internal audit should help not only to support the strategic planning, but also to identify the areas of improvement as it can be seen on the left side of the model. There is not often a procedure or method to integrate this information, specifically for the business improvement activity. In the words of some managers, they commented “…We don’t have at the moment a procedure, but the self-assessment process takes about three months and then it is reviewed by consultants/examiners so as they can draw conclusions…” another manager said: “…We write an application for the quality award every year. [We produce a document which is updated every year], [we collect all useful information from production, the finances, etc.] Everything that is required to produce the document, for the application of the quality award…”

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure5.png}
\caption{Managing information to support strategic planning and business improvements.}
\end{figure}

Literature reviews suggest there is paucity of methods not only to support continuous improvement, but also to follow up other activities after self-assessment. Quality Foundations [5], p.18 and [18], p.25, provide guidelines for managing information and knowledge. However, the frameworks do not provide specific approaches for doing this, and more importantly, none of the frameworks had considered the use of knowledge and information to support systematically strategic planning and business improvements activities. As a result, there is not only a lack of specific methodologies for managing knowledge in self-assessments, but also a lack of structured approaches to integrate the knowledge into a business improvement strategy. In fact, the methods in which the EFQM model supports business improvement strategies and strategic planning are unclear and seemed not to be formalized.

4. DEVELOPING A BUSINESS IMPROVEMENT STRATEGY

Developing an effective business improvement strategy is not an easy task for any organization, particularly when there are important barriers in terms of resources such as people, capital and technology. These factors are crucial and significantly determine the success or failure of quality programs implementations, consequently, these factors

\textsuperscript{5} Surveys, questionnaires, and reports from internal and external audits.
should be carefully considered before implementing a business improvement strategy. Hammer and Champy [19] were right when they foresaw the important role of information technology, capital, people and processes that organizations were going to have. Today, it is a reality, and unfortunately, for the majority of companies and their directors, these resources are limited and scarce. Thus, the developing of an improvement business strategy should contain in the right proportion and with the adequate quality the factors mentioned above, and failure to get them right will result in pitfalls or inadequate implementations.

The framework proposed in this paper to develop a business improvement strategy is based on the deployment of the EFQM model derived from the emerging issues of an empirical study with twelve European organizations. This framework, however, represents a case of best practice in which organizations have found a good way to use self-assessments based on the EFQM model. It does not necessarily represent a general way of using the EFQM model, neither a prescriptive recipe to integrate self-assessments with current business strategies. In addition, this framework needs further developments and requires to be tested either through modeling techniques or with empirical implementations. Thus, the method summarizes some best practices and intends to serve as a guideline to use self-assessments outcomes to construct or modify current business improvement strategies.

4.1 SELF-ASSESSMENTS OUTCOMES AND BUSINESS IMPROVEMENT STRATEGY

In this way, this approach is subject to the current situation of the organization along with its internal and external environments. If an organization or directors decides to employ this framework, it has to be led by top management and the relevant strategy/quality management departments, as well as with the support of external consultancy if needed. It also requires detailed planning for every stage involved that need to be tailored based on particular needs, culture, and the availability of resources. This framework can be deployed using the following steps:

1 Diagnostic organizational situation: determining the needs

The first step is to conduct a diagnostic of the organization and an analysis of the external environment. This will help to determine the needs and the role and objective of deploying the EFQM model. This decision should be linked to current strategic goals, that is, the requirements in terms of business results in getting from point A to point B. The radar methodology suggested by the European Foundation for Quality Management addressed at the begging of this paper is a good start as suggested in [20].

2 Tailor and complement the EFQM model criteria

It might be necessary to tailor the EFQM model criteria considering geographical location, government regulations, product/service issues, culture, industry in which the businesses are in, etc. These factors should be considered carefully to assure that model criteria are relevant to the organization. It is also very important to consider the actual maturity level of quality that organizations have so that they select the correct tools and techniques and set realistic goals.

3 Deploy the EFQM model using self-assessments

This process is concerned with the deployment of the EFQM model through a series of logical steps as suggested in [2] and [20]. Some authors, consultants, and quality foundations provide comprehensive support for this stage, so, it is recommended that organizations make the best of it. Since this is a critical stage, it is recommended that organizations seek advice and assistance.

4 Conduct internal/external business intelligence (BI)

First, this stage refers to the task of looking for and selecting several sources of information that can provide key parameters/data in benchmarking, industry tendencies, financial/economic data/facts, product/service demands, and country intelligence. The accuracy and relevance of this information will depend on organizations’ needs, and specific requirements to functional and divisional areas (e.g., marketing department, financial department, quality department, product/service divisions, etc.). It is very important to allocate the appropriate resources to look for the data and transform it into business knowledge, as key information is not usually free. The organization should also consider the adequate infrastructure in terms of information technology and the qualified human resources in the required disciples to construct this framework.
Developing a Business Improvement Strategy based on self-assessments using the EFQM Model

Second, there should be considered the structure of internal indicators that can help to support development and innovation of product and services. There are some valuable resources of information that can help organizations to issues of benchmarking, case studies, best practices, etc. Other sources of internal analysis may come from self-assessments outcomes and feedback.

5 Analyze self-assessments and BI

This step refers to the analysis and discussion of self-assessments and business intelligence and should take place at a business strategy level. It is necessary that people involved know very well the business improvement and strategy agendas, so that they make the best decisions. Since the top management makes final decisions, they must know the information in detail to support the decision-making with objective analyses of self-assessments and BI.

6 Construct the business improvement strategy agenda

This is the factual formulation of a detailed strategy and action plans to improve key processes at operational levels. It is also concerned with the selection of quality improvement initiatives (e.g., lean, six sigma, BPR, ISO 9000 series, statistical process control, etc.) that tackle the specific issues emerged from stage 6.

This process should be included in the business strategy agenda and be monitored at all times to ensure that business improvement objectives are met. The full process should also be synchronized with self-assessments and business strategy reviews as suggested in [11].

5. SUMMARY

This paper examined the integration of self-assessments using the EFQM model to develop a business improvement strategy. It presented the core issues of the EFQM model and addressed the difficulties that

---

Consider business databases provided by EBSCO Source premier, Emerald management Xtra, Sage management publications, ABI Inform, among others. For Business Intelligence consider Bureau van Dijk data services, Compustat, Global insight, Global market information database, Global best practice, excellence one, and a business performance resources such as BPIR.
organizations have had to follow up self-assessment outcomes. It then explored the current linking issues between the self-assessment and business improvements, emphasizing the role of business strategy to support this issue.

The paper is based on some emerging issues of a study with twelve European organizations that used the EFQM model for several years. Based on this insights gathered through semi-structured interviews, the authors argue that it is necessary and feasible to support effectively the development of business improvement strategies based on self-assessment activities. Then, a method is proposed to effectively integrate and follow up self-assessments and to develop a business improvement strategy, setting also the agenda for the improvement plans and activities. The proposed framework also addressed the integration of business intelligence as a core stage to construct business knowledge, and to help to speed up decision-making.

It identified the need to conduct further research to investigate the way in which the knowledge generated with improvement activities and self-assessments can be managed/integrated through current knowledge management frameworks. Finally, the paper pointed out the necessity to standardize the integration process self-assessments both, with the agenda of business improvement strategies and the strategic planning process.

REFERENCES


Application of Quality Function Deployment to Study Critical Service Quality Characteristics and Performance Measures

Baba Md Deros¹, Norashikin Rahman, Mohd Nizam Ab. Rahman, and Ahmed Husam Said

Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, Malaysia.

Abstract

This paper aims to measure service quality performance and identify critical service quality characteristics as perceived by the customers. This began by conducting a survey through designed questionnaires to the identified group of customers to get their feedback. This is followed by constructing house of quality, establishing the customer’s requirements and technical specifications, design the questionnaire, and identifying other parameters for the Quality Function Deployment (QFD) matrix. There are two different competitive assessments that have been conducted; performance related to customer requirements and performance related to technical requirements. Both case study companies are involved in the same industry with studied firm. Based on the QFD technique, critical service quality characteristics that need to be enhanced by the studied firm in Malaysia have been identified as technical specification such as employee training, human resource, customer feedback analysis, understanding customer needs, quality and implementation of continuous improvement, process management, and supply from suppliers. From the study, it was found the service quality performance of the studied firm was 68 percent, which the firm’s top management feels unsatisfactory. Therefore, the authors had suggested to the studied firm to further improves its service quality performance by improving the critical service quality characteristics as perceived by its customers.

1. Introduction

In today’s global and highly competitive market, it is essential for the survival of any firm involved in the service industry to be adaptive, responsive to changes, proactive and has the capability to deliver high quality products according to diverse customer requirements. Therefore, it is very important for any firms which involved in service industry to improve their service quality by reducing the gap between internal quality and external customer satisfaction [1]. The main objectives of this study are to measure the service quality performance and identify critical service quality characteristics as perceived by the firm’s customers. QFD is a customer driven planning process to guide the design, manufacturing, and marketing of goods. In this study, QFD will be used as a tool to improve quality in service industry. By implementing QFD, it will help the firms involved to have clearer picture of quality requirements that could improve their customers’ satisfaction. QFD is a powerful technique in determining customers’ requirements and integrate them to the product or service provided by system integrator firm. QFD technique is not only applicable in manufacturing industry, but it could be implemented in various industries to meet or exceed customer expectations. By using QFD technique, service quality characteristics and performance of studied firms can be determined.

2. Literature Review

QFD is one of the TQM quantitative tools and techniques that could be used to translate customer requirements and specifications into appropriate technical or service requirement. This is important in order to deliver product or service that fulfills or exceeds customer requirements. According to Guinta and Praizler [2], QFD is a customer driven tool. While Chan and Wu [3], stated that QFD is the customer driven planning process that could guide the

¹ Corresponding author: Tel.: (603)-8921 6117; Fax: (603)-8925 9659; E-mail: hjbaba@eng.ukm.my
design, manufacturing and marketing of goods. QFD uses visual planning matrices that link customer requirements, design requirements, target values and competitive performance into one chart [4].

QFD was first put into use at Mitsubishi’s Kobe shipyard site in 1972, and later in 1983 it was introduced into the USA [5]. In 1966, Yoji Akao had introduced the QFD concept in Japan [6]. Professor Mizuno first used QFD in 1972 to Mitsubishi’s Kobe shipyard site to design super tankers [7]. According to Cohen [8], the two pioneer researchers that had developed the QFD were Mizuno and Akao. In the late 1970s, Toyota the automobile manufacturer had adopted QFD and further developed the QFD concept to a detailed process [8]. According to Sullivan [9], Toyota Auto body had started using QFD in 1977, and as a result, Toyota introduced four new van-type vehicles between 1977 and 1984. Toyota experienced a 20% reduction in the startup cost due to QFD technique adoption in launching its new products from 1977 till 1979 [3].

In 1986, Ford Motor Company and Xerox were the early users of QFD that initiated the use of QFD concept in the USA [3]. Since then, QFD has been developed and broadly used in various industries such as automotives, electronics, banking, insurance, healthcare, utilities, food processing, aerospace, software engineering, construction and marketing [10]. Many other multinational companies such as IBM, HP, General Motors, AT&T, Digital Equipment, ITT, Baxter Healthcare, Texas Instruments, Miliken Textile, Black and Decker and Philips International have subscribed to QFD advantages [11]. The Cadilllac car model 1992, considered one of the great car models that had attracted many customers at that time, this car model has been planned and designed entirely using the QFD technique [12].

3. Methodology

Case study methodology was used in this research. It starts from identifying specific group of customers that the firm deals with in order to get their feedback (voice of customer) on the designed case study protocol. Case study protocol was designed to give detailed feedback needed for making improvements to the provided service or product quality. This case study protocol includes information about the importance weight of customer requirements, service performance, competitive assessment, and customer suggestions (open ended questions) on the customer requirements field. The authors will then integrate these voices of customers into the service or product provided by incorporating these collected data into QFD matrix (house of quality). For constructing the QFD house of quality, it involves establishing the primary inputs to QFD matrix such as customer requirements (WHATs), technical specifications (HOWs). The relationship matrix between (WHATs) and (HOWs), and correlation between (HOWs) will also be established. Other QFD matrix parameters that shall be included are the importance weight of customers’ requirement, competitive assessment, relative and absolute weight of customers’ requirements and technical specifications, technical difficulties and target. Establishing customer requirements is the most important step in QFD process. In this research, the main customer requirements (WHATs) have been adapted from Garvin [13]; Parasuraman et al. [14] and modified by the firm’s quality team and experts. Second, it is important to establish technical specifications in QFD process. In order to meet customers’ requirements, the firm’s quality team and experts had identified the technical specifications (HOWs). Other parameters for QFD matrix had also been identified. They were assessed using the case study protocol; the relative and absolute weights of the customer requirements and technical specifications were calculated from QFD equations, the firm’s quality team assesses the technical difficulties and target. To achieve research goals, service quality performance and critical service quality characteristics that need to be enhanced are defined during data analysis and evaluation. The data collected from case study protocol has been integrated and analyzed to the QFD matrix. Then, the critical technical specifications were identified. These shall have the highest relative and absolute weights that the firm should focus on and improve, in order to meet or exceed customer expectations and improve the service quality performance provided by the firm. The last activity was measuring the service quality performance of the case study firms. Finally, the quantitative value in % of the firm’s quality performance was measured by dividing the actual performance level to the maximum performance level.
4. RESULTS AND DISCUSSIONS

4.1 COMPETITIVE ASSESSMENTS

As mentioned earlier, the two main aims of this study are to measure service quality performance and identify critical service characteristics that need to be enhanced in order to improve the studied firm’s customer satisfaction. This can be achieved after the completion of constructing QFD matrix.

![Figure 1: Identify critical service characteristics](image)

Table 1: General firm background comparisons between studied firm and competitors A and B

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Studied firm (P)</th>
<th>Competitor A</th>
<th>Competitor B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid Up Capital</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Over Head</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Number of Staff</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Resources</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Product Reputation</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Ability to Finance Large Projects</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Nation Wide Recognition</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>
Table 1 represents general background information with respect to firm size between the studied firm, P, and the other two competitors, A and B, that could provide better understanding about the size of three firms assessed. There are seven dimensions that authors had considered to classify the firms into small, medium or large size. With reference to Table 1, the authors had classified the studied firm (P) as medium size firm, while competitor A as large size firm and competitor B as medium size firm.

<table>
<thead>
<tr>
<th>Management involvement</th>
<th>Customer focus and public</th>
<th>Technology implementation</th>
<th>Process management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encourage and inherent the quality responsibility to all of the firm staff.</td>
<td>Customer feed back and analysis</td>
<td>Increase the Warranty period</td>
<td>Create customer data base</td>
</tr>
<tr>
<td>Quality and continuous improvement implementation</td>
<td>Understand the customer needs</td>
<td>Supply from suppliers</td>
<td>Information technology and Web collaboration</td>
</tr>
<tr>
<td>The degree of the top management involvement, encouragement, empowerment, and support</td>
<td>Knowledge about the market and profitable products</td>
<td>Human resources</td>
<td>Resources</td>
</tr>
<tr>
<td>Understanding customer relation and satisfaction</td>
<td>Increase the Warranty period</td>
<td>Employee training</td>
<td>Process management</td>
</tr>
</tbody>
</table>

Figure 2: Technical specifications competitive assessment

Referring to Figure 2, it is shown that the overall performance of the studied firm (P) is better than competitor (B) related to technical specifications. On the other hand, the overall performance of competitor (A) is better than the studied firm (P) related to the same technical specifications.

The studied firm (P) is performing better than competitor A in some technical requirements, which can be considered as the firm’s strengths. Technical specifications have been divided into twelve groups as shown in Figure 3. Since P has high expertise in this field, which provides system integration of building management systems, access control system, public address system, intercom system, and other sub system, it is performing better in terms of knowledge about the market and profitable products. Other than that, P is performing better in customers’ relation and satisfaction. These capabilities had made P to perform better in terms of customer focus and public relations. Compared to competitor A, which had already achieve the status by being a well established firm, P performs better in participating in seminars and exhibitions to reach its target to be a well known firm. This is important because P needs to advertise itself and to make known to the market. The studied firm (P), is performing better in term of dealership and product warranty period which the studied firm has extend it to three years rather than one year compared to other competitors. P is a dealership from several high quality products supplied by recognized worldwide suppliers for its products and systems provides better confidence to P to increase the warranty period.

Studied firm (P) is achieving lower than competitor A in some of the technical requirements, which can be considered as the firm’s weaknesses. Lack of employee training and resources lead to poor service quality and process management. Small number of staff had caused the firm to perform lower than competitor A in term of human resource. Since the firm does not have enough human resources, it depends on the suppliers or contractors to perform the system installation. As a result, it makes difficult for the firm to control the quality of installed system. In terms of management involvement, technology implementation and understanding customers’ need, getting supply from suppliers and resources, P is performing lower than A. These specifications include the degree of top management involvement, encouragement, empowerment and support, implementation of quality and continuous
improvement as well as encouragement of quality responsibility to all the firm staff. All these specifications need to be improved to ensure customers’ satisfaction to the delivered products and services. Continuous improvements and high quality products will lead P to be recognized by customers.

In order for P to stay competitive in today’s market, there are some improvements need to be implemented. Majority of the firm’s staff are managerial level, therefore the firm has to recruit more technical and engineering staff in order to improve the performance of system installation and customer services. Increasing the firm’s technical staff will help the studied firm (P) to reduce its dependence on the contractors and have better quality control on the executed projects. Proper technical and customer service training shall be provided to the newly hired staff. P also needs to develop a standard specifications and criteria for suppliers’ selection for better performance, quality control and process management. Contractors and suppliers that perform system installation shall be selected based on these standard criteria so that the quality of installed system could be monitored and always at the satisfied level. P also needs to understand more on customer needs to achieve better performance than other competitors. In terms of management involvement, all specifications need improvements especially in quality and continuous improvement implementations as well as to encourage quality responsibility to all the firm staff. More programs and training on quality awareness shall be held to promote the importance of quality in the products and service. High quality products will sustain the customers and improve the firm reputation. Quick response time and critical analysis to every customer feedback are important for P to give a better service, thus could increase the firm’s status.

Referring to Figure 3, it also shows that the overall performance of the studied firm (P) better than competitor B related to the customer requirements. However, overall performance of competitor A is slightly better that the studied firm related to the same customer requirements.

There are twenty eight requirements which have been divided into seventeen main groups, as shown in Figure 3. Studied firm, (P), is performing better in terms of conformance by meeting customers’ specifications for the system installed including the workmanship. For product consistency and reliability, P is performing better by providing consistent of system/ product performance over time. In terms of accessibility, P is performing better compared to other competitors because P has friendly and accessible staff in solving any problems. Other than that, customers are easy to access to the firm’s staff in order to obtain contracting service. P which is capable to focus on meeting customers’ requirements performs better in terms of responsiveness. For communication requirement, P has the ability to listen to the customer complaints and suggestions, which lead to better performance compared to other competitors.

Beside all these strengths, studied firm (P) has to improve its performance related to the customer requirements, which is lagging behind the two competitors (A and B). Low performance related to minimum time to complete the projects, actual project duration, completeness and responsiveness to the projects can be considered as the firm’s weaknesses. These can be improved by enhancing the human resource, employee training, customers’ feedback analysis, customer understanding, quality and continuous improvement and enhancing the process management.

It is important for P to improve its performance in some customer’s requirements, so that the reputation and the firm’s status could be increased and better from the other two competitors. In terms of service, P need to hire more technical staff to provide prompt resolution of system/ product problems and complaints as well as good service and technical support during the warranty period. P also could benefit to these technical staffs as it could minimize the completion time and improve responsiveness to the variation of requirements during completing the project. Since P is performing at the average level between other two competitors in terms of accuracy and precision, P needs to develop standard procedure to the staffs to ensure they perform the right service from the first time with the minimum errors to the products/systems installed. Small good changes to every aspect could improve the firm’s performance and minimize the cost.
Figure 3: Customer requirements competitive assessment
4.2. SERVICE QUALITY PERFORMANCE

Service quality performance for the studied firm is calculated in terms of quantitative value by dividing the total maximum service performance level to the total actual service performance level. The author had followed the following steps to calculate the service quality performance. These steps have been adapted and modified from Arditi and Lee [15]:

1. Actual performance is identified from the questionnaires administered to the firm’s customers.
2. Maximum performance is the maximum performance the firm could perform which is excellent and ranked with numerical number (5).
3. Actual service performance = \[\text{final importance weight of each customer requirements} \times \sum \text{relationship value between WHAT and HOW for the technical requirements} \times \text{actual performance}\] (Equation 1)
4. Maximum service performance = \[\text{final importance weight of each customer requirements} \times \sum \text{relationship value between WHAT and HOW for the technical requirements} \times \text{maximum performance}\] (Equation 2)
5. Total actual service performance level = \[\sum \text{Actual service performance}\] (Equation 3)
6. Total maximum service performance level = \[\sum \text{Maximum service performance}\] (Equation 4)
7. Service quality performance = \[\frac{\text{Total actual service performance level}}{\text{Total maximum service performance level}}\] (Equation 5)

From case study protocol analysis and competitive assessments for both, customers’ requirements and technical specifications and following steps provided by Arditi and Lee (2003), the service quality performance can be calculated by using Equation (5):

\[
\text{Service quality performance of the studied firm} = \frac{20505.23}{30282.45} = 0.67713
\]

\[
= 67.713\%.
\]

The result from the calculation shows the current quality service performance of the studied firm (P) is about 68%. Therefore, it is very critical for the studied firm to further improve its service quality performance to higher level in order to ensure their customers are satisfied, thus provide long term sustainability and growth.

After determining the service quality performance and identifying the critical service characteristics that need to be improved, the firm should have clearer vision of its strengths, weaknesses, opportunities and threats (SWOT). In other words, the firm had benchmarked itself against its competitors and identified the critical areas that need to be enhanced in order to improve customer satisfaction.

6. CONCLUSION

QFD is an effective technique that helps both manufacturing and non-manufacturing industries to improve their quality, performance, customer satisfaction, understanding the customers’ needs, benchmarking against competitors, and clear vision of customer, market intangible requirements, and other quality and business characteristics by integrating the voice of customer with the firm’s processes. Based on the QFD technique used in this study, critical service quality characteristics that need to be enhanced by the studied system integrator firm in Malaysia have been identified, and service quality performance of the studied firm has been determined and quantified.

The critical service quality characteristics that need to be enhanced by the firm are the technical specification which includes employee training, human resource, customer feedback analysis, understanding customer needs, quality and continues improvement implementation, process management, and supply from suppliers. The authors could conclude that the studied firm (P) is capable of installing good quality products but they do not have enough competent staff to install the products. To overcome this problem, the studied firm (P) is using contractors for the installation of the systems that might not be able to provide installation quality up to the customers’ expectation.
Therefore, the studied firm (P) should ensure and enhance the supply from suppliers, which are the contractors for installing the system by awarding the installation job only to high quality installers.

The calculated service quality performance of the studied firm is 68%, which is considered not satisfactory to the firm’s top management. Therefore the studied firm (P) should further improve its service quality performance to higher level by improving the critical service quality characteristics as required by their customers.

Since most previous researches on service quality had focused on the qualitative quality characteristics. Therefore more researches that focus on the quantitative quality characteristics should be conducted in the future. More studies are needed to quantify service quality using QFD techniques to show that QFD technique can also be applied in service industry and not only limited to manufacturing industry.

This study had achieved its aims to identify the critical service quality characteristics and determine the service quality performance of a system integrator firm in Malaysia using QFD technique. In future, the QFD technique could be used by the studied firm (P) to get the importance weight of the customer requirements from customers that the firm had not dealt with them earlier. This is to identify the critical quality characteristics that need to be enhanced by the firm in order to enhance the firm’s ability to attract more customers, get more business and market shares, thus able to achieve better customer satisfaction.

7. ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Science Technology and Environment (MOSTE) Malaysia and Universiti Kebangsaan Malaysia for their support in providing the research grant for the project entitled “Development of a tool for benchmarking implementation in manufacturing SMEs” (Science Fund 06-01-02-SF0345).

REFERENCES

Manufacturing Process Improvement; Perceived Barriers from Adoption of Good Practice by SMEs.

Smith P.G\textsuperscript{1}, Oduoza C.F.\textsuperscript{1}, and Barber K\textsuperscript{2}.

\textsuperscript{1}School of Engineering & the Built Environment  
University of Wolverhampton  
Shifnal Road, Priorslee, Telford, TF2 9NT, UK

\textsuperscript{2}School of Management  
Bradford University  
Emm Lane, Bradford, BD9 4JL, UK

ABSTRACT

Several waste minimisation techniques that can lead to manufacturing process improvements are available to small and medium sized enterprises, however, a recent study suggests that there is a gap between theoretically researched knowledge base and its uptake especially by small and medium size businesses. An understanding of the barriers that potentially prevent SMEs from not using waste minimisation techniques may help to explain the reasons why the uptake is low and thereby enable them to gain the benefits from abundant knowledge provided by researchers over the years.

This paper aims to identify the barriers and constraints facing small and medium size businesses thus preventing them from implementing waste minimisation techniques resulting from academic research. Future research will aim to develop a suitable tool especially for small and medium size enterprises that could support waste minimisation and thereby boost productivity within their businesses.

1. INTRODUCTION

The current economic climate is one of uncertainty. Therefore, it is now even more important to look internally at businesses to develop the capability to minimise business waste and therefore improve performance. For some businesses it may be a simple task to just set some new objectives for the improvement department to “get on with it”. Though many businesses in industry may have the skills and resources there may be barriers that prevent such improvements ever taking place.

Evidence suggests that some organisations are no more than 30% effective, the key message here is not to let a lack of paper qualifications become an automatic barrier to problem solving, as sometimes is the case, Macadam\textsuperscript{[7]}.

There is a need to research the potential barriers that small businesses face in initiating projects and minimising waste.

2. PERCEIVED GAP OF NOT IMPLEMENTING CONTINUOUS IMPROVEMENT TECHNIQUES

Although there is a wealth of information available to SMEs from various sources for example: academic citings from Journals and books, Google and the Internet, or consultants. It has been shown that the uptake from SMEs of CI techniques is low Oduoza \textit{et al.}\textsuperscript{[11]}. Mulhaney, Sheehan and Hughes\textsuperscript{[10]} have identified the problems that SMEs often face: they do not have the resources of larger organisations, no one person responsible for driving improvements, cost, effort, administration, lack of awareness, using external consultants may not generate sustainable benefits, lack of external funding (particularly in the current economic climate), where to obtain external funding and the fear of change. The concept “fear of change” is also supported by Dwyer and Copland\textsuperscript{[4]} and McQuater \textit{et al.}\textsuperscript{[9]}. Dwyer and Copland\textsuperscript{[4]} in his case study work demonstrates that CI puts too much pressure on already-stretched resources and also fear of the unknown leads to a lack of commitment to see things through, this is also supported by Harris\textsuperscript{[6]}. Cutbacks by organisations in training and education budgets are also being implemented as immediate cost saving measures as a reaction to the economic downturn. Small businesses have to make the best use of resources in terms of; money, people, facilities and technology to ensure that they are used effectively and efficiently. Small manufacturers typically find themselves short of technical and financial resources and any cutting of training and development budgets can only make this situation worse.
McQuater et al. [9] highlight the problems of SMEs in implementing process improvements as being attributable to lack of management support and also a lack of understanding of the process. Lack of education into CI techniques is also supported by, Arndt [1]. Failure for most change processes can be traced to strategy not being linked to business goals, Price and Chahal [12], Sanders [14] and Recardo [13]. Esain [5] when researching the use of Value Stream Mapping (VSM) highlighted the extent of training required to make use of the tools indicated that it could be as long as 15 days for each person.

Delmas and Toffel [3] studied organisational responses to environmental demands and noted the influence of social and cultural influences on strategic decisions made by organisations. They propose that differences in the influence of corporate departments lead their facilities to prioritise different pressures and thus adopt different management practices. It is strongly believed that organisations facing common institutional pressures may decide to adopt different management practices depending on prioritisation and preferences for specific operations management approaches. It is possible therefore, that a potential barrier to adoption of good practice is the choice especially by SMEs to adopt quick fix approaches rather than tried and tested established tools and techniques which would normally provide optimal solutions.

3. POTENTIAL BARRIERS TO TECHNOLOGY TRANSFER TO BUSINESSES– A LITERATURE REVIEW

The General Accounting Office [17] in the United States has criticised American Universities for their enthusiasm in developing new technologies and not moving them into private sector applications. Similarly, policymakers insist that the time lag between the discovery of new knowledge at the university and its exploitation by businesses could seriously impair the global competitiveness of firms. Marshall [8] to eliminate any obstacles and facilitate adoption of new technologies in a university / industry technology transfer (UITT), governments passed the Patent and Trademark act removing any restrictions on licensing and allowing universities to own and manage intellectual property. The key objective was to accelerate the commercialisation of new technologies in order to promote adoption of new technologies and promote economic development and entrepreneurial success. It is thought that in spite of efforts made by universities to foster technological diffusion, there is still a perceived gap between theory and practice, Oduoza et al. [11].

Seigel et al. [16] have researched barriers preventing the transfer of scientific knowledge to practitioners in industry and an abridged data from their study is summarised in Table 1. It shows data with regards to the technology transfer barriers perceived by managers, TTO directors and university scientists. According to their study, industry appears to lack in understanding of what academic institutions and academics do. Academics also are thought to be out of touch with businesses particularly the needs of small businesses with extreme resource constraints. According to table 1, the figures represent the percentage of respondents for a particular item of barrier to University / Industry Technology Transfer.

Table 1: Barriers to University/Industry Technology Transfer, Seigel et al. [16].

<table>
<thead>
<tr>
<th>Barriers to Technology Transfer</th>
<th>Managers (%)</th>
<th>TTO Directors /Administrators (%)</th>
<th>University Scientists (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of understanding regarding university corporate or scientific norms and environments</td>
<td>90.0</td>
<td>93.3</td>
<td>75</td>
</tr>
<tr>
<td>Insufficient rewards for university researchers</td>
<td>31.6</td>
<td>60.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Bureaucracy and inflexibility of university researchers</td>
<td>80.0</td>
<td>6.6</td>
<td>70.0</td>
</tr>
<tr>
<td>Insufficient resources devoted to technology transfers by universities</td>
<td>31.6</td>
<td>53.3</td>
<td>20.0</td>
</tr>
<tr>
<td>Poor marketing /technical/negotiation skills of TTOs</td>
<td>55.0</td>
<td>13.3</td>
<td>25.0</td>
</tr>
<tr>
<td>University too aggressive in exercising intellectual property rights</td>
<td>80.0</td>
<td>13.3</td>
<td>25.0</td>
</tr>
<tr>
<td>Faculty members/administration have unrealistic expectations regarding the value of their technologies</td>
<td>25.0</td>
<td>40.0</td>
<td>10.0</td>
</tr>
<tr>
<td>‘Public domain’ mentality of universities</td>
<td>40.0</td>
<td>8.3</td>
<td>5.0</td>
</tr>
<tr>
<td>Number of interviews</td>
<td>20</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>
Table 2 shows the suggested improvements in the reduction of barriers in the process of technology transfer according to Siegel et al. [16]. According to table 2, the figures represent the percentage of respondents who identified a particular item as a suggestion for improvement to University / Industry Technology Transfer. The suggestions made are aimed to address the barriers to adoption as specified in Table 1. Overall, the suggestions clearly reflect the feelings of respondents about how to facilitate University / Industry Technology Transfer.

**Table 2: Suggested Improvements to University Technology Transfer Process Seigel et al. [16].**

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Managers (%)</th>
<th>TTO Directors/ Administrators (%)</th>
<th>University Scientists (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universities and industry should devote more efforts to developing better mutual understanding</td>
<td>80.0</td>
<td>93.3</td>
<td>75.0</td>
</tr>
<tr>
<td>Modify reward systems to reward technology transfer activities</td>
<td>85.0</td>
<td>80.0</td>
<td>80.0</td>
</tr>
<tr>
<td>Universities need to provide more education to overcome information and cultural barriers</td>
<td>85.0</td>
<td>86.7</td>
<td>60.0</td>
</tr>
<tr>
<td>Universities should devote additional resources to technology transfer</td>
<td>45.0</td>
<td>46.7</td>
<td>60.0</td>
</tr>
<tr>
<td>Universities should be less aggressive in exercising intellectual property rights</td>
<td>55.0</td>
<td>10.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Increase networking between scientists and practitioners</td>
<td>35.0</td>
<td>26.7</td>
<td>40.0</td>
</tr>
<tr>
<td>Universities need greater marketing expertise in the TTO</td>
<td>50.0</td>
<td>20.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Number of interviews</td>
<td>20</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

Cooper and Edgett [2] have studied how to overcome the crunch in resources for new product development and demonstrated the negative impact of resource constraints on projects and conclude that business goals are the starting point in which to improve. Their model described in figure 1 shows that several underlying variables such as poor people morale, lack of resource, insufficient time and expertise, waiting for people’s actions and too many projects with limited resource are key to poor project management and adoption of innovative ideas.

![Figure 1: The negative impact of resource constrains on projects (Adapted from Cooper and Edgett [2].)](image)

4. RESEARCH METHODOLOGY

The methodologies used in this project are a combination of literature review to obtain current information produced by academics on barriers to SMEs when minimising waste and semi-structured telephone interviews
conducted with 25 senior management employees in SMEs across the UK. Table 3 provides a summary of the responses from the interview.

5. RESULTS

Data from table 3 shows that of the 25 companies interviewed, 60% do not have any form of business goals to minimise waste. A total of 76% did not contractually have to implement CI techniques as part of their customer requirements. Poor technical knowledge of processes was cited by 40% of respondents and this lack of understanding of processes will inevitably contribute to business waste. In total, 72% of companies did not measure business waste and therefore had no real control over it. When interviewed about any CI techniques used, none of the respondents had any knowledge of the available techniques or any business improvements achieved in their organisations. When interviewed about any barriers for not using CI techniques 20 companies could not state any reasons while others complained about time constraint, pressure due to assigned production targets, cultural differences, investment issues and changes to normal working procedures. 48% of interviewees complained about lack of top management commitment. When asked about the potential barriers to adopting continuous improvement tools, interviewees attributed the difficulties in the following proportions; production pressure to supply customers 46%, lack of support 30%, investment 14% and training 8%. The last question explored links with universities, however not one person mentioned them.

When people were interviewed about waste minimisation their views were very blinkered solely to packaging waste. All of the people interviewed did not mention any CI techniques cited by academics or consultants and often discussed in trade journals.

Table 3: Telephone Interviews held with some SMEs in the UK.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMEs that do not have business goals.</td>
<td>60%</td>
</tr>
<tr>
<td>CI techniques not required by customer contract.</td>
<td>76%</td>
</tr>
<tr>
<td>Poor technical understanding of business processes.</td>
<td>40%</td>
</tr>
<tr>
<td>Business waste not measured.</td>
<td>72%</td>
</tr>
<tr>
<td>Top 5 techniques used to minimise waste?</td>
<td>0%</td>
</tr>
<tr>
<td>Reported successes from using techniques</td>
<td>0%</td>
</tr>
<tr>
<td>SMEs that are not sure of barriers preventing improvements</td>
<td>Companies - 20</td>
</tr>
<tr>
<td>Top management commitment</td>
<td>62%</td>
</tr>
<tr>
<td>Awareness of specific issues with reducing waste</td>
<td></td>
</tr>
<tr>
<td>(a) Production</td>
<td>46%</td>
</tr>
<tr>
<td>(b) Support</td>
<td>30%</td>
</tr>
<tr>
<td>(c) Finance</td>
<td>14%</td>
</tr>
<tr>
<td>(d) Training</td>
<td>8%</td>
</tr>
<tr>
<td>SMEs identified unaware where to seek additional improvement support.</td>
<td>Not sure - 64%</td>
</tr>
<tr>
<td>Internet</td>
<td>16%</td>
</tr>
<tr>
<td>Regional Development Agency</td>
<td>4%</td>
</tr>
<tr>
<td>Government</td>
<td>4%</td>
</tr>
<tr>
<td>Consultant</td>
<td>4%</td>
</tr>
<tr>
<td>Council</td>
<td>4%</td>
</tr>
<tr>
<td>Department of Trade</td>
<td>4%</td>
</tr>
</tbody>
</table>

6. DISCUSSION

There is abundant evidence of perceived barriers from adoption of good manufacturing practice by SMEs [2,3,17, 8, 10, 11 and 16]. For instance in their model for the effective transfer of scientific knowledge from academicians to practitioners, Siegel et al.[16] based on structured interviews of various university/industry technology transfer (UITT) stakeholders conclude that there are numerous impediments to effectiveness in UITT. Typical issues are cultural and informational barriers among stakeholder types and administrators,
academics and firms / entrepreneurs, inadequate rewards for involvement, and sometimes process complexity which reduces personnel morale and decision to circumvent the process.

Another inhibitor to adoption of continuous improvement especially by SMEs is the lack of resources available to them and the fact that they are in competition with much larger organisations where resources are not as constrained. Such constraints have been reported as barriers to the implementation of ISO 9001 in SMEs. Wilson [15] reported that cost and effort have constituted serious impediments to SMEs acquiring and maintaining ISO 9001 certification. They viewed it as consuming a lot of administrative effort without a potential for long term business improvement. The implication is that awards for business excellence for SMEs become more difficult to achieve. Cooper and Edgett [2] have echoed the issue of resource constraint to process improvement in their paper, overcoming the crunch in resources for new product development. They recommend adequate product portfolio management and tactical solutions tracking the deployment of existing resources.

Dwyer and Copland [4], Nissan struggles with continuous improvement identifying the barriers to adopting continuous improvement tools as: employees attitude, poor problem definition and wrong perception of what CI tools are supposed to be doing; consequently it becomes difficult to match a typical problem with the right CI tool. Therefore increasing an unwillingness to accept change.

Companies identified in this publication do tend to value interaction with universities, however the problems still exist. Once the barriers are resolved progress can be made. Considerations to improve are: reward system for UITT, review of staffing practices, flexible university policies, additional resources, eliminate cultural issues, ref Seigel et al. [16].

Our research study based on literature review and a preliminary survey has established that there are barriers to technology adoption by SMEs. Key barriers to adoption are limited resources, complexity of new technology, poor top management commitment, lack of adequate training, mismatch of problem definition with available tools and unwillingness to change. However not all issues documented by academics were raised in the interviews to SMEs. SMEs are still not sure in regards of the largest issues in terms of not using CI techniques to minimise waste.

If SMEs understood more about CI techniques and the potential benefits and this was transferred effectively by universities would the uptake by SMEs increase? There is still much research to be conducted to minimise the issues that have been discovered.

7. CONCLUSION

The study carried out here has highlighted how continuous improvement tools and techniques could enhance company productivity and business performance. Unfortunately, literature review and preliminary survey shows that businesses are failing to adopt these continuous improvement tools for a variety of reasons ranging from; lack of awareness to lack of resources and sometimes low preference / prioritisation for elaborate technology / methodology for business process improvement possibly due to poor understanding of the application of these tools. There is therefore the need to develop a suitable user-friendly model / framework especially for SMEs that could support waste minimisation and facilitate enterprise business processes.

8. FUTURE WORK

To conduct further research to determine if SMEs have the desire to improve and if the benefits were made clear to them would it increase the uptake? Also to research why SMEs only believe waste minimisation is relevant only to the reduction of packaging. Research what can universities do differently to help SMEs improve the uptake of CI techniques.

9. REFERENCES


Modeling Thermal Radiative Properties of Nano scale Multilayer with Incoherent Formulation

S.A.A.Oloomi
Department of Material Engineering, Islamic Azad University, Yazd Branch, Yazd, Iran
amirooloomi@iauyazd.ac.ir

ABSTRACT

Radiative properties of a material are the core of thermal science and optics, which play critical roles in modern technologies, including MEMS/NEMS. The radiative properties, such as reflectance, transmittance, and emittance of multilayer structures largely depend on the direction and wavelength of incident radiation as well as wafer temperature. They are also affected by thin-film coatings and surface roughness [1].

The studied examples using silicon wafer and either silicon dioxide or silicon nitride coating demonstrate the strong influence of coating and coating thickness on the radiative properties. This study helps gain a better understanding of the radiative properties of semitransparent wafers with different coatings and will have an impact not only on semiconductor processing but also on thin film solar cells.

Fud studied radiative properties of NIMs by using three multilayer structures with NIM layer [1]. We have interested in studying nanoscale radiative properties with silicon because silicon is the most extensively used material in MEMS/NEMS.

1. INTRODUCTION

Radiative properties of a material are the core of thermal science and optics, which play critical roles in modern technologies, including MEMS/NEMS. The radiative properties, such as reflectance, transmittance, and emittance of multilayer structures largely depend on the direction and wavelength of incident radiation as well as wafer temperature. They are also affected by thin-film coatings and surface roughness [1].

The studied examples using silicon wafer and either silicon dioxide or silicon nitride coating demonstrate the strong influence of coating and coating thickness on the radiative properties. This study helps gain a better understanding of the radiative properties of semitransparent wafers with different coatings and will have an impact not only on semiconductor processing but also on thin film solar cells.

Fud studied radiative properties of NIMs by using three multilayer structures with NIM layer [1]. We have interested in studying nanoscale radiative properties with silicon because silicon is the most extensively used material in MEMS/NEMS.

During the past two decades, there have been tremendous developments in near-field imaging and local probing techniques. Examples are the Scanning Tunneling Microscope (STM), Atomic Force Microscope (AFM), Near-field Scanning Optical Microscope (NSOM), Photon Scanning Tunneling Microscope (PSTM), and Scanning Thermal Microscope (SThM).

The ability to manufacture, control, and manipulate structures at extremely small scales is the hallmark of modern technologies, which include microelectronics, MEMS/NEMS, and nanobiotechnology. Spectral and directional control
of thermal radiation is a challenging yet important task for a number of applications, such as thermophotovoltaic (TPV) energy conversion, solar energy utilization, space thermal management, and high-efficiency incandescent lamps.

Pattern-induced radiative property variations can be an important problem for the wafer temperature measurement and the temperature uniformity control during integrated circuit manufacturing. In addition, light diffraction can be used to monitor the etching depth and other features during the microfabrication and lithographic processes.

Heat transfer at nanoscale may differ significantly from that in macroscales. With device or structure characteristic length scales becoming comparable to the mean free path and wavelength of heat carriers (electrons, photons, phonons, and molecules), classical laws are no longer valid and new approaches must be taken to predict heat transfer at nanoscale.

Rapid thermal processing (RTP) has become a key technology for semiconductor device manufacturing in a variety of applications, such as thermal oxidation, annealing, and thin-film growth.

Temperature measurements and control are critically important for continuous improvement of RTP [2, 3]. Since the heating source is at a much higher temperature than that of the silicon wafer, radiative energy exchange is the dominant mode of heat transfer. Hence, understanding the radiative properties of silicon and other relevant materials is essential for the analysis of the thermal transport processes. Furthermore, since many RTP furnaces use noncontact lightpipe thermometers, accurate determination of the wafer emittance is necessary for correlating the radiance temperature to the true wafer temperature [4, 5]

This Paper, predict the directional, spectral, and temperature dependence of the radiative properties for the multilayer structures consisting of silicon and related materials such as silicon dioxide, and silicon nitride.

2. MODELING

2.1. COHERENT FORMULATION

When the thickness of each layer is comparable or less than the wavelength of electromagnetic waves, the wave interference effects inside each layer become important to correctly predict the radiative properties of multilayer structure of thin films. The transfer-matrix method provides a convenient way to calculate the radiative properties of multilayer structures of thin films (Figure 1).

Figure 1: The geometry for calculating the radiative properties of a multilayer structure

By assuming that the electric field in the $j$th medium is a summation of forward and backward waves in the $z$-direction, the electric field in each layer can be expressed by

$$E_j = \begin{cases} A_j e^{iq_j z} + B_j e^{-iq_j z} e^{i(q_j z - toff)}, & j = 1 \\ A_j e^{iq_j z} + B_j e^{-iq_j z} e^{i(q_j z - toff)}, & j = 2, 3, \ldots, N \end{cases}$$

(1)

Where $A_j$ and $B_j$ are the amplitudes of forward and backward waves in the $j$th layer. Detailed descriptions of how to solve for $A_j$ and $B_j$ is given in [1].
2.2. Incoherent Formulation

When the thickness of silicon substrate is much greater than the coherent length, and the considered wavelength falls in the semitransparent region of silicon, interferences in the substrate are generally not observable from the measurements. In this case, the incoherent formulation or geometric optics should be used to predict the radiative properties of the silicon substrate. Two ways to get around this problem are to use the fringe-averaged radiative properties and to treat thin-film coatings as coherent but the substrate as incoherent [1] (Figure 2).

![Schematic of thin-film coatings on both sides of a thick silicon](image)

Figure 2: Schematic of thin-film coatings on both sides of a thick silicon

Figure 2 shows the geometry of the silicon wafer with thin-film coatings on both sides. Note that $\rho_{is}$ and $\tau_{i}$ are the reflectance and transmittance, respectively, of the multilayer structure at the top surface (air-coatings-silicon) for rays incident from air, assuming that the silicon extends to infinite. On the other hand, $\rho_{is}$ and $\tau_{b}$ are for rays incident from silicon. Note that the transmittance $\tau_{i}$ is the same when absorption inside silicon is negligibly small [6]. Similarly, $\rho_{bs}$ and $\tau_{b}$ are for the multilayer structure at the bottom surface for rays incident from the substrate. The transfer-matrix method can be separately applied to calculate the reflectance and transmittance at the top and bottom surfaces of the wafer, by neglecting the absorption of silicon.

The absorption of silicon can be taken into consideration by introducing the internal transmittance

$$\tau_{i} = \exp\left(-\frac{4\pi k_{s} d_{s}}{\lambda \cos \theta_{s}}\right)$$  \hspace{1cm} (2)

Here, $k_{s}$ is the extinction coefficient of silicon, $d_{s}$ is the thickness, and $\theta_{s}$ is the angle of refraction. The angle of refraction is complex due to absorption.

For a slightly absorbing medium with $k_{s} \ll 1$, however, $\theta_{s}$ can be determined using Snell’s law by neglecting absorption [7]. Consequently, the radiative properties of the silicon wafer with thin-film coatings in the semitransparent region can be expressed as [8]

$$\rho = \rho_{is} + \frac{\tau_{i}^{2} \rho_{bs}}{1 - \tau_{i}^{2} \rho_{is} \rho_{bs}}$$ \hspace{1cm} (3)

$$\tau = \frac{\tau_{i} \tau_{b} \tau_{h}}{1 - \tau_{i}^{2} \rho_{is} \rho_{bs}}$$ \hspace{1cm} (4)

$$\varepsilon = 1 - \rho - \tau$$ \hspace{1cm} (5)

3. Optical Constants

The optical constants, including the refractive index ($n$) and the extinction coefficient ($k$), of a material are complicated functions of the wavelength and temperature.
They also depend on the crystalline structure as well as doping and impurity levels. In the present work, carefully selected empirical expressions are used to calculate the optical constants of lightly doped silicon (doping concentration \(< 10^{15} \text{ cm}^{-3}\)).

3.1. The Refractive Index of Silicon

Jellison and Modine [10] measured the ratio of the Fresnel reflection coefficients of silicon wafers in both polarization states with a two-channel spectroscopic ellipsometer in the temperature range from 25 °C to 490 °C. From the measurement results, they extracted the refractive index and extinction coefficient using the least-squares Levenberg-Marquardt fitting.

The Jellison and Modine (J-M) expression of the refractive index for wavelength between 0.4 μm and 0.84 μm is given by

\[ n_{JM}(\lambda, T) = n_0(\lambda) + \beta(\lambda)T \]  \hspace{1cm} (6)

\[ n_0 = \sqrt{4.565 + \frac{97.3}{3.648^2 - (1.24/\lambda)^2}} \]  \hspace{1cm} (7)

\[ \beta(\lambda) = -1.864 \times 10^{-4} + \frac{5.394 \times 10^{-3}}{3.648^2 - (1.24/\lambda)^2} \]  \hspace{1cm} (8)

where \( \lambda[\mu m] \) is the wavelength in vacuum and \( T[°C] \) is the temperature.

Li [3] extensively reviewed the refractive index of silicon. By carefully analyzing published experimental data, he developed a functional relation, based on the modified Sell Meier type dispersion relation, for the refractive index of silicon that covers the wavelength region between 1.2 μm and 14 μm and the temperature range up to 480 °C.

\[ n_L(\lambda, T) = \sqrt{\varepsilon_\infty(T) + \frac{g(T)\eta(T)}{\lambda^2}} \]  \hspace{1cm} (9)

\[ \varepsilon_\infty(T) = 11.631 + 1.0268 \times 10^{-3}T + 1.0384 \times 10^{-6}T^2 - 8.1347 \times 10^{-10}T^3 \]  \hspace{1cm} (10)

\[ g(T) = 1.0204 + 4.8011 \times 10^{-4}T + 7.3835 \times 10^{-8}T^2 \]  \hspace{1cm} (11)

\[ \eta(T) = \exp(1.786 \times 10^{-4} - 8.526 \times 10^{-6}T - 4.685 \times 10^{-9}T^2 + 1.363 \times 10^{-12}T^3) \]  \hspace{1cm} (12)

To calculate the refractive index of silicon, in this study the J-M expression is used in the wavelength region from 0.5 μm to 0.84 μm, and Li’s expression at wavelengths above 1.2 μm. In the wavelength range between 0.84 μm and 1.2 μm, we use a weighted average based on the extrapolation of the two expressions.

Notice that beyond 6 μm or so, lattice vibration causes additional absorption; however, its effects can be neglected due to the weakness of the phonon oscillator in silicon. Figure 3 shows the calculated refractive index of silicon in the wavelength range from 0.5 μm to 10.0 μm at selected temperatures. In general, the refractive index of silicon decreases as the wavelength increases and the temperature decreases.

When 10.0 \( \lambda > \mu m \), the refractive index of Si is assumed to be independent of the wavelength, and the value of \( n_L(\lambda = 10, T) \) is used to represent the refractive index of Si at any temperature \( T \) for \( \lambda > 10.0 \) μm.
3.2. The Extinction Coefficient of Silicon

The extinction coefficient (\(k\)) and absorption coefficient (\(\alpha\)) are related by \(\alpha = \frac{4\pi k}{\lambda}\).

The absorption coefficient of silicon depends on the absorption processes, such as interband transition, intraband transition, and free-carrier absorption. When the photon energy is higher than the band gap energy of silicon, electrons in the valance band can be excited to the conduction band, resulting in a large absorption coefficient.

The J-M expression of the extinction coefficient, covering the wavelength range from 0.4 \(\mu\)m to 0.84 \(\mu\)m, is given as [10]:

\[
k_{JM}(\lambda, T) = k_0(\lambda) \exp \left[ \frac{T}{369.9 - \exp(-12.92 + 6.831/\lambda)} \right]
\]

\[
k_0(\lambda) = -0.0805 + \exp \left[ -3.1893 + \frac{7.946}{3.648^2 - (1.24/\lambda)^2} \right]
\]

The absorption coefficient can be deduced from the extinction coefficient.

In the longer wavelength region, Timans [8] measured the emission spectra of several silicon wafers and deduced the absorption coefficient in the wavelength region from 1.1 \(\mu\)m to 1.6 \(\mu\)m, in the temperature range between 330°C and 800°C. He suggested that the absorption coefficient can be expressed as a summation of the band gap absorption and free-carrier absorption as following:

\[
\alpha(\lambda, T) = \alpha_{BG}(\lambda, T) + \alpha_{FC}(\lambda, T)
\]

The expression for the band gap absorption can be found in the work by MacFalane et al. [11] and is given by

\[
\alpha_{BG}(\lambda, T) = \sum_{i=1}^{4} \alpha_{a,j}(\lambda, T) + \sum_{i=1}^{2} \alpha_{e,i}(\lambda, T)
\]

(14)

Notice that silicon is an indirect-gap semiconductor and the absorption process is accompanied by either the absorption of a phonon, denoted by \( \alpha_{a} \), \( (\lambda, T) \), or the emission of a phonon, \( \alpha_{i} \) denoted by \( \alpha_{e} \), \( (\lambda, T) \). Detailed expressions for \( \alpha \) \( (\lambda, T) \) and \( \alpha_{e} \) \( (\lambda, T) \) can be found in Timans [8, 12].

The band-gap absorption disappears at wavelengths longer than that corresponding to the energy gap (modified by the phonon energy).

For the free-carrier absorption, Sturm and Reaves [13] suggested an expression based on their measurement of the transmission of the wafer at 1.30 \( \mu m \) and 1.55 \( \mu m \) and in the temperature range of 500 \(^{\circ}C\) to 800 \(^{\circ}C\). The Sturm and Reaves (S-R) expression is

\[
\alpha_{FC} = N_e A_e + N_h A_h
\]

(15)

Where \( N_e \) and \( N_h \) are electron and hole concentrations, and \( A_e \) and \( A_h \) are electron and hole absorption cross sections, respectively. The S-R expression agrees well with experimental results in the wavelength region between 1.0 \( \mu m \) and 1.5 \( \mu m \), but departs from experiments at longer wavelengths. Vandenabeele and Maex [14] studied the free-carrier absorption of silicon in the infrared region by measuring the emission from the double-side-polished silicon wafers at the wavelengths of 1.7 \( \mu m \) and 3.4 \( \mu m \), in the temperature range from 400 \(^{\circ}C\) to 700 \(^{\circ}C\). They proposed a semi-empirical relation for calculating the extinction coefficient as functions of wavelength and temperature due to free-carrier absorption.

The Vandenabeele and Maex (V-M) expression is given by

\[
\alpha_{FC}(\lambda, T) = 4.15 \times 10^{-5} \lambda^{0.51} (T + 273.15)^{2.95} \times \exp\left(\frac{-7000}{T + 273.15}\right)
\]

(16)

Here again, \( T \) is in \(^{\circ}C\). Rogne et al. [15] demonstrated that the absorption coefficient calculated from the V-M expression agrees well with experimental data in the wavelength region between 1.0 \( \mu m \) and 9.0 \( \mu m \) at elevated temperatures.

In the wavelength region between 6.0 \( \mu m \) and 25.0 \( \mu m \), lattice vibrations causes an additional absorption. Since the effect of lattice absorption is negligible in most RTP applications compared to the absorption by free carriers, it is assumed to be independent of the temperature and dopant concentration. The extinction coefficient for lightly doped silicon due to the lattice absorption is simply obtained from the tabulated extinction coefficient values given in Ref. [16] at room temperature.

The absorption coefficient of lightly doped silicon is determined from the JM expression at \( \lambda < 0.9 \mu m \) and the Timans expression is combined with V-M expression for the free-carrier absorption at \( \lambda > 0.9 \mu m \).

Figure 3 shows the calculated extinction coefficient of silicon in the wavelength range from 0.5 \( \mu m \) to 9.0 \( \mu m \) at selected temperatures. Notice that unlike the refractive index of silicon, there exists a discontinuity in the calculated extinction coefficient especially at the elevated temperatures because the functional expression is changed at \( \lambda = 0.9 \mu m \). The existence of the absorption edge is clearly seen from the figure, and the extinction coefficient generally increases as the temperature increases. Since the V-M expression is applicable up to 9.0 \( \mu m \), the empirical models for the extinction coefficient of lightly doped silicon are inevitably limited to the wavelengths less than 9.0 \( \mu m \).
Figure 4: Calculated extinction coefficient of silicon at selected temperatures

The optical constants of silicon dioxide and silicon nitride are mainly based on the data collected in Palik’s handbook [17, 18] and are assumed to be independent of temperature.

4. Results

Consider the case in which the silicon wafer is coated with a silicon dioxide layer on both sides. The thickness of silicon wafer is \(500\ \mu\text{m}\) and the temperature of silicon wafer with thin-film coatings is 25°C, and the EM waves are incident at \(\theta = 0\). The considered wavelength range is \(0.5\ \mu\text{m} < \lambda < 2.0\ \mu\text{m}\). Some results of this study are shown below in figures 5 to 13.

Figure 5: Spectral Emittance of silicon wafer coated with silicon dioxide film with different thicknesses (200 nm, 400nm, 600nm, 800nm) on both sides, at room Temperatures and normal incidence with Incoherent Formulation.
Figure 6: Spectral Transmittance of silicon wafer coated with silicon dioxide film with different thicknesses (200 nm, 400nm, 600nm, 800nm) on both sides, at room Temperatures and normal incidence with Incoherent Formulation.

Figure 7: Spectral Reflectance of silicon wafer coated with silicon dioxide film with different thicknesses (200 nm, 400nm, 600nm, 800nm) on both sides, at room Temperatures and normal incidence with Incoherent Formulation.
Figure 8: Spectral Emittance of silicon wafer coated with silicon dioxide film (400 nm thickness) on both sides at different Temperatures (25°C, 500°C, 1000°C) with Incoherent Formulation.

Figure 9: Spectral Transmittance of silicon wafer coated with silicon dioxide film (400 nm thickness) on both sides at different Temperatures (25°C, 500°C, 1000°C) with Incoherent Formulation.
Figure 10: Spectral Reflectance of silicon wafer coated with silicon dioxide film (400 nm thickness) on both sides at different Temperatures (25°C, 500°C, 1000°C) with Incoherent Formulation.

Figure 11: Spectral Emittance of silicon wafer coated with different materials (silicon dioxide and silicon nitride) on both sides (400 nm thickness), at room Temperatures and normal incidence with Incoherent Formulation.
Figure 12: Spectral Transmittance of silicon wafer coated with different materials (silicon dioxide and silicon nitride) on both sides (400 nm thickness), at room Temperatures and normal incidence with Incoherent Formulation.

Figure 13: Spectral Reflectance of silicon wafer coated with different materials (silicon dioxide and silicon nitride) on both sides (400 nm thickness), at room Temperatures and normal incidence with Incoherent Formulation.

5. Conclusions

We have analyzed and calculated the spectral, directional and temperature dependency of radiative properties of a three layers material using transfer-matrix method. In the present work, carefully selected empirical expressions for calculating the optical constants of materials.

The layer thicknesses need to be optimized to achieve maximum transmittance for the given materials. Maximum transmittance also depends on the type of materials and its temperature.
For lower wavelengths, more emittance occurs in thicker coatings. The reflectance decreases as the temperature increases, because of increasing emittance.

Silicon dioxide coating has higher emittance than silicon nitride coating for $0.7\mu m < \lambda < 0.9\mu m$, but for $0.55\mu m < \lambda < 0.7\mu m$ and $0.9\mu m < \lambda < 1.0\mu m$ silicon nitride coating has higher emittance than silicon oxide coating. Therefore coatings act as wavelength selective emitters for radiative energy conversion and thermal radiation detection.

6. References


Simulation of service-oriented and distributed manufacturing systems

Nylund Hasse*, Salminen Kai, and Andersson Paul H

Department of Production Engineering
Tampere University of Technology
Tampere, 33101, Finland

ABSTRACT

In this paper modeling and simulation is discussed in the context of distributed manufacturing systems (DiMS). The DiMS concept aims towards more efficient and innovative collaboration that integrates the design and development activities of manufacturing systems. A DiMS, described formally as a digital manufacturing system, is a competent basis for the modeling and simulation of manufacturing systems. The activities of the simulation model are described on the basis of a service-oriented approach in which the content of the services is completely known. The micro, meso, and macro manufacturing levels are applied to modeling and simulation. By means of the levels the hierarchy of the communication of the services is illustrated. An example is described in which modeling and simulation is used to proceed from ideas to efficiently operated manufacturing systems.

1. INTRODUCTION

The motivation for this paper is the change towards networked collaboration, which is caused by, for example, globally distributed markets, the potential of information and communication technology (ICT), and the specialization of manufacturing companies in their core competences. It has led to increasingly complex manufacturing activities in the manufacturing network, and the importance of cooperation has become a critical factor. The Distributed Manufacturing Systems (DiMS) approach strives towards an integrated environment focusing on the complexity of manufacturing activities and their connections. It explains the system by autonomous and co-operative entities, capable of fulfilling their individual activities and of collaborating in a dynamically changing environment. The complexity of manufacturing systems requires information and knowledge to be managed digitally. The need emerges for a formal and standard way to present the information and knowledge that, with a service-oriented approach, can lead to simplified information and knowledge management. An essential part of digitally presented manufacturing systems is the use of modeling and simulation thorough the process from ideas to solutions. By using modeling and simulation, the principles by which solutions are reached can be analyzed and compared. On the basis of the analysis, suitable alternative solutions can be selected for detailed examination prior to the actual realization. This can lead towards more lean and sustainable solutions, reducing the non-value-adding activities and increasing the reuse of the information and knowledge in the collaboration.

2. FMS 2010 CONCEPT OF DISTRIBUTED MANUFACTURING SYSTEMS

The DiMS concept aims to enable collaborative and autonomous manufacturing systems to respond more efficiently to challenges in dynamically changing global markets. The goal is to integrate the design and development activities of products, production systems, and business processes, supporting innovativeness from ideas and visions to efficient results. The DiMS concept is realized in the form of service-oriented and learning-based activities connecting requirements with existing capabilities and future possibilities. The information and knowledge of manufacturing systems is stored digitally, which makes it possible to have up-to-date information and knowledge that evolves with each manufacturing activity. This enables the accuracy of the information and knowledge to be used as the basis of future design and development activities.

* Corresponding author: Tel.: (+358) 50-5746050; Fax: (+358) 3-31152753; E-mail: hasse.nylund@tut.fi
2.1. Structure of Distributed Manufacturing Systems

The basic building blocks of DiMS are entities of products, resources, and orders, as well as their related manufacturing domains of process, production, and business [1]. The entities, despite being involved in different manufacturing activities, have generally similar structures. Similarly, the manufacturing domains have different roles, and the combined collaborative activities of the entities and domains make up the total manufacturing system. Figure 1 shows the structure of the manufacturing system and the system entities. The manufacturing entities are autonomous and co-operative as they collaborate within the same manufacturing system. Products represent what is offered to customers and resources are static entities of a manufacturing system, for example, machines, tools, and devices. Orders as instances of products, the dynamic entities of a manufacturing system that appear in the system, flow through the system and disappear. The presentation of the manufacturing entities is based on the Holonic Manufacturing Systems (HMS) reference architecture Product-Resource-Order-Staff Architecture (PROSA) [2]. The entities are connected with the process, production, and business domains. The process domain offers the manufacturing capabilities and the production domain uses the resources efficiently to manufacture the right orders at the right times. The business domain is responsible for markets, i.e. for the right products being available for customers.

![Figure 1: Structure of DiMS entities and activities, as well as internal structure of the entities [3].](image)

The entities have internal structures consisting of digital, virtual, and real parts. The digital part contains the information and knowledge which can be presented digitally, whilst the real part represents what exists in the real manufacturing system. The virtual parts usually present the physical parts as computer models. While the digital, virtual, and real parts present the autonomous nature of a manufacturing entity, the communication part is the capability of the entity to collaborate in the system together with other entities. The Fractal Manufacturing System (FrMS) approach explains the self-similarity of manufacturing entities at different structuring levels [4]. A manufacturing system is a part of some larger system and at the same time consists of subsystems. A manufacturing system is also a part of a supply chain, its horizontal context co-operating with e.g. suppliers, partners, and customers.

3. Digital Manufacturing System Integrating Manufacturing Activities

The research related to digital manufacturing systems has no commonly specified definitions. However, there are shared similar descriptions, which include, for example: collaborative and distributed networks for product design [5]; the integration of product development and production planning [6]; an effort to develop a solution to integrate all the applications and data present in the manufacturing environment [7], and using information technology and computer simulation to model real-world manufacturing processes in order to analyze and understand them [8].

In this paper the digital manufacturing system is defined as “an integrated environment for the design and development of products, production systems, and business processes” [9]. A digitally presented manufacturing system contains the information and knowledge of manufacturing entities and activities that it is reasonable to represent in a digital form. This, at its best, enables there to be efficient collaboration between all manufacturing
activities and related parties. A digital manufacturing system also includes computer-aided tools for the simulation, modeling, and analysis of manufacturing systems. Examples of characteristics of digital manufacturing systems are:

- Integrated platform for manufacturing design and development activities.
- Digital, virtual, and real existence of manufacturing systems.
- Time concepts of past, present, and future.

Design and development activities can be integrated into one digitally presented manufacturing system. The requirements of design activities can be viewed against existing manufacturing capabilities, and the need for new capabilities can be determined. Developing a manufacturing system adds or changes the characteristics of existing requirements or capabilities, and the updated state of the system is instantly available for future tasks.

A digital manufacturing system in a wide sense can be divided into digital, virtual, and real domains of existence, which co-exist in an extended digitally presented system, following the proposed structure of manufacturing entities. The information and knowledge of the manufacturing system exists in the digital portion. It exists only once and the real and virtual portions can both access it. When a real manufacturing system is in operation, it gets the information and operation rules that it needs from the digital domain and updates the data when the state of the real system changes. Similarly, the virtual portion, such as a simulation model, uses the same digital information and knowledge.

The digital manufacturing system exists in the dimensions of past, present, and future, each of them having a different purpose. The past exists as digital information and knowledge collected from the manufacturing activities. In the present system, in the time dimension of now, the digital and real systems co-exist. The digital system is used to aid the decisions to be made in the real system, and the real system also logs the manufacturing activities into the digital system, which is constantly getting new information. In the future dimension the digital, virtual, and real parts co-operate supporting visions and strategies for future activities.

3.1. MANUFACTURING SYSTEM AS A COMPLEX SET OF KNOWLEDGE-INTENSIVE SERVICES

A manufacturing system can be seen as multiple autonomous manufacturing entities interacting and co-operating in a complex network of manufacturing activities. The activities are explained as services, which hold the information and knowledge needed to explain the manufacturing activities. It is required that the activities are known exactly, in that they are understood by all related parties. Describing the activities as services in a digital format creates a formal way to present the services. This makes possible efficient collaboration in a digital manufacturing system between entities that can be humans, machines, or information systems. The information and knowledge is kept as the autonomous property of the manufacturing entities and the communication between the entities includes only the information that is needed to fully describe the collaboration activity.

The collaboration in the digital manufacturing system is twofold, including the management of information and knowledge, as well as managing communications. The knowledge management of manufacturing systems consists of four essential processes: creation, storage and retrieval, transfer, and application, which are dynamically changing and continuous phenomena [10] as each manufacturing activity changes the content and meaning of the existing information and knowledge. The communication between the manufacturing entities is based on service-oriented architecture (SOA), which consists of self-describing components that support the composition of distributed applications [11] enabling the autonomous manufacturing entities to negotiate and share their information and knowledge. The basic conceptual model of the SOA architecture consists of service providers, service requesters, and service brokers [12]. Table 1 explains the roles of manufacturing entities in a digital manufacturing system based on SOA.

<table>
<thead>
<tr>
<th>Roles</th>
<th>Description of the actor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service requester</td>
<td>Typically the product entities when they are realized as order entities. The order entities call on the services they require to be manufactured.</td>
</tr>
<tr>
<td>Service provider</td>
<td>Usually the resource entities which have the capabilities needed to provide the services that are requested.</td>
</tr>
<tr>
<td>Service broker</td>
<td>An actor that contains the rules and logics of using the services. Its function is to find service providers for the requesters on the basis of criteria such as cost, quality, and time.</td>
</tr>
</tbody>
</table>
Figure 2: An example of a service in detail; the hierarchy of the service is divided into multiple services on a lower level [13].

Figure 2 shows an example of a service occurring in the process domain between products and resources. It also illustrates how a service on an upper level is divided repetitively into multiple services on the lower level to the level of part features as requesters and manufacturing methods as providers. The service consists of roles, context, and receipt. A service consists of two different entities, i.e. the product and resource entities having the roles of service requesters and providers. The actual service, being the manufacturing activity, is twofold, consisting of a context and receipt. The context is the environment, real or virtual, where the service takes place, whilst the receipt is the digital description of the service. The division into roles, context, and receipt is based on the distributed character of knowledge in social practices, where interactive situations are the context, positions the role, and dispositions represent the receipt [14].

The product entity requests a service, which is provided by the resource entity. The service, whether it is happening in a virtual or real environment, has a certain context that is in a certain state. The state is a basis for the actions happening during the service, and the result is based on the skills of the service provider. During the service both resource and product entities collect data from the process. The collected data are analyzed, forming information that is the basis for learning from the service. When something is learned, it is used to update the receipt, which will be the basis for future services. When a certain product entity uses a service provided by a certain resource entity, the data collection, analysis, learning, and updating phases include adding the same data and information to the knowledge of both entities. The knowledge of a resource entity is updated with several product entities using the services it provides. In a similar fashion, the knowledge of a product entity consists of all the services it requests. A service can be seen as a hierarchy in which a service on the upmost level divides iteratively into multiple sub-services until the level on which the individual part features are requested. This means that an entity requesting a service gets information about the possible service provider entities, but it does not know how the service request is fulfilled. For example, a service request for the manufacturing of a product is a request on the macro level. The macro level service request is divided into multiple sub-services on the meso level and the meso-level service makes similar requests on the micro level. The upper level only needs the information about whether the service request can be fulfilled or not. The hierarchy of the services may be limited by the service requester as it may state special requirements for the service that limit the selection of possible providers. For example, a customer may require certain parts of ordered products to be manufactured in a specific manufacturing plant.

4. MODELING AND SIMULATION AS A PART OF DIGITAL MANUFACTURING SYSTEMS

The modeling and simulation of manufacturing systems is one of the areas in which simulation is most widely used. It is a powerful method to design new manufacturing systems and to develop existing ones to improve their performance [15]. A model can be explained as a representation of an object, i.e. a system or phenomenon, without the object itself being used. Modeling and simulation is used, for example, to understand the behavior of the object to achieve a specific result, to find out what happened and why it happened. It can also be used to repeat simulation experiments to refine the performance of the object, to formalize the object in order to apply it to a different content or context, and to derive inspiration for something new that is loosely based on the simulated object [16]. The following sections explain the levels of modeling and simulation, the use of simulation from ideas to solutions, and benefits and challenges connected with modeling and simulation in the context of digital manufacturing systems.
4.1. MICRO, MESO, AND MACRO LEVELS IN MODELING AND SIMULATION

The terms macro and micro have been used in the area of human resource management (HRM), where the macro area focuses on organizations and the micro area focuses on individuals or small work groups with shared identities [17]. Similarly, in economics, the macro level is an analysis of complex structures and associated processes, while the micro level is about individual carriers of rules and the systems they organize. A meso level can be added between the macro and micro levels. It is a population that is made up of individuals on the micro level and it is also an element in the higher-order level, the macro level [18]. The modeling and simulation of manufacturing on the system level is usually referred to as a macro-level simulation in which the behavior of a whole manufacturing plant is examined. The micro level is used to research individual manufacturing activities, such as machining processes. This leaves a gap, as the macro system cannot be explained with micro-level activities. Applying the meso level to the context of modeling and simulation enables the macro and micro levels to be connected. A meso-level study focuses on how a group of micro-level activities act in a macro-level environment. Table 2 briefly illustrates the micro, meso, and macro levels in the domain of modeling and simulation.

Table 2: General descriptions of micro, meso, and macro levels of modeling and simulation.

<table>
<thead>
<tr>
<th>Modeling and simulation level</th>
<th>Description of the level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Micro level</strong></td>
<td>Manufacturing units representing individual machine tools, manufacturing methods, and work pieces. Areas of modeling are, for example, computer-aided design (CAD) and computer-aided manufacturing (CAM).</td>
</tr>
<tr>
<td><strong>Meso level</strong></td>
<td>Manufacturing stages, physical or logical areas, consisting of manufacturing units and their co-operation. Typical examples are the designing of cell layout and material handling and holding, as well as human resource simulation.</td>
</tr>
<tr>
<td><strong>Macro level</strong></td>
<td>Manufacturing system composed of manufacturing stages, as well as material and information flow between the manufacturing stages. Simulation is used in, for example, factory and layout design, as well as in measuring the performance of the system.</td>
</tr>
</tbody>
</table>

Figure 3 presents the macro, meso, and micro levels of modeling and simulation. Inputs for the macro levels are customer orders, which consist of changing volume and variation of parts from one or more part families, and the outputs are, for example, delivery times and the reliability of customer orders and the utilization versus capacity of the system. A manufacturing stage represents a meso level, which uses workers and handles material within a manufacturing stage. The output information is tact times and descriptions of manufacturing capabilities. A machining unit is an example of a micro-level feature, which can be viewed as an individual process in which one specific part feature or more is machined using certain machining methods. The output on a micro level is the quality of the work, as well as knowledge of the possible methods that can be used.

Figure 3: Inputs, outputs, and interaction between different levels of manufacturing modeling and simulation.
The interaction between the levels is implemented in the form of services. A service request on the macro level is a customer order. The macro-level model sends a service request to the meso level. The customer order is divided into individual parts, and the service request is a batch order of one or more parts. The batch order can be a combination of multiple macro-level requests, depending on the nature of the meso-level manufacturing stage. The response from the meso level may include, for example, work times, how long the service will take, and the cost of the service. If there is more than one possible provider on the meso level, the decision as to which one will be selected is taken at the macro level for each service request independently. The service from the meso to the micro level is divided in a similar fashion into multiple sub-services. The service request is for a work piece including the required features to be manufactured. The response will include the quality and duration of the machining process. Similarly to what happens on the macro level, the provider is decided on the meso level if multiple possibilities exist.

The micro, meso, and macro levels can also be focused differently on the fractal structuring levels. A manufacturing system, as a micro-level entity, is a part of a supply chain, a meso-level entity. The supply chain is formed of entities and operates in an industrial network, which in this approach is the macro level. A similar approach is, for example, a macro-institutional perspective on supply chains, describing the levels as a firm as a micro, the supply chain as a meso, and the region of operation as a macro feature [19].

4.2. USING MODELING AND SIMULATION FROM IDEAS TO SUCCESSFUL SOLUTIONS

Modeling and simulation can be used in many different areas of designing and developing manufacturing systems. Figure 4 presents an example that progresses from ideas and visions to an efficiently operating system. The process is incremental, consisting of multiple phases which will draw closer to the solution. It is also iterative as it is possible to go back to previous phases in order to refine them. Each of the phases has inputs and outputs in addition to the results from the previous phase. As the process evolves, the inputs and outputs will have more precise digital and virtual presentations. The example has both digital and virtual existences but it can have connections to the real world too so as to make sure that the solutions can be realized when the virtual existence is applied to a real system.

![Diagram](image)

Figure 4: Use of modeling and simulation in the process from the need for change to successful solutions.

The need for change can arise from external or internal factors. Examples of external factors are the economy, the environment, or social and political factors [20], while the internal factors are ideas and visions of the future of the manufacturing system, based on strategic decisions and customer expectations. These, viewed against the capabilities within the existing manufacturing system, form the new requirements for the future system. The new requirements are
solutions. The oblique effects on other entities also have to be evaluated before decisions are made.

Conceptual stages. For efficient collaboration, all the manufacturing activities between the entities and their related domains have to be considered. Optimizing one area without taking the context into account may lead to deficient results. The modeling and simulation can be in different phases of the process, which enables manufacturing entities to be evaluated as early as in their conceptual stages of design and development activities. In the process that moves from ideas and visions to efficient solutions, the modeled entities of a manufacturing system can be investigated, which results in alternative solutions. The solution alternatives can be modeled as virtual entities that include, in addition to their digital description, for example, 3D models with their operating rules, motion, and behavior. The virtual entities need to have corresponding equivalents existing as real manufacturing entities in the manufacturing ecosystem, e.g. as suppliers and partners, or they have to be implemented. Combining the existing and new virtual entities forms the first version of a simulation model in which the entities perform their own tasks and cooperate with other entities. The solution that is implemented has to be verified to make sure that the behavior and cooperation of the entities in the system are modeled correctly. The verified simulation model evolves to the state where test simulation runs can be executed using test experiments of which the outcome is either known or could be predicted by analyzing the possible outcome in detail before the test simulation runs. By analyzing the results from the simulation model and comparing them with the predicted outcome, the behavior of the simulation model can be validated. When the simulation model is verified and validated, it can be used for manufacturing experiments. The experiments can lead to an efficiently operating manufacturing system, the behavior of which is fully known. This can lead towards a successful solution.

On the basis of the principles of Extreme Programming (XP) [21], a simulation model of entities from different phases of the proposed process may be constructed. For example, a macro-level simulation will get the work times from a meso-level simulation and the meso-level entity can be anything from a digitally described entity to a fully working simulation model. If the meso-level entity is in the digital entity phase, it will return the duration of the work without any calculations. The value could be based on collected history data or the desired target value of a future process. At the other end, the fully working model on the meso level may simulate the process and return the duration as a simulation result. The benefit is that the behavior of the macro-level system can be investigated during the early stages of design and development activities, and the process from ideas to solutions of the meso-level entities can be modeled within the context where they will exist in the future.

5. Conclusions

Being involved in the construction of a digitally presented manufacturing system increases the understanding of the behavior of the system. The experts on different manufacturing activities acquire a wider outlook compared to their special domain of knowledge. Simulation models provide an efficient way to demonstrate events that are happening in the manufacturing system. The effects of individual decisions on the whole system are easier to understand by using real-looking 3D models and animations of the system and the experiments do not disturb the real manufacturing system in operation. Storing the history of the system as digital information and knowledge provides an opportunity to look back and find out what happened and what caused it to happen. It is possible to find the root causes and prevent or minimize unwanted events from happening in the future.

A digital manufacturing system relies heavily on information and knowledge. This can be divided into explicit and tacit knowledge [22]. Explicit knowledge can be presented with symbols which make it possible to present it formally and digitally, while tacit knowledge consists of human beliefs, know-how, and skills. The challenge is how to represent and maintain tacit knowledge [10]. The accuracy of the digitally presented information and knowledge has to be constantly evaluated, with possible faults being eliminated when the information is updated.

Describing a manufacturing system on the basis of the integrated DiMS architecture in terms of digital, virtual and real existences in past, present, and future time creates an efficient platform for the design and development activities of manufacturing systems. Presenting the information and knowledge as services simplifies the knowledge management as only the information that is needed is transferred between the manufacturing entities and the knowledge is kept as their autonomous property. The services can be applied to descriptive modeling and simulation models with macro, meso, and micro levels that communicate through services. Modeling and simulation can be used in a process that moves from ideas and visions to efficient solutions. The modeled entities of a manufacturing system can be in different phases of the process, which enables manufacturing entities to be evaluated as early as in their conceptual stages. For efficient collaboration, all the manufacturing activities between the entities and their related domains have to be considered. Optimizing one area without taking the context into account may lead to deficient solutions. The oblique effects on other entities also have to be evaluated before decisions are made.
ACKNOWLEDGEMENTS

The FMS 2010 research project is part of the SISU 2010 Innovative Manufacture technology programme. It is co-financed by Tekes, the Finnish Funding Agency for Technology and Innovation, and several major companies in Finland. It is a joint effort of several laboratories within the Department of Production Engineering (DPE) of Tampere University of Technology (TUT).

REFERENCES


CLASSIFICATION, CODING AND RETRIEVAL OF MECHANICALLY ORIENTED PRODUCTS USING CUSTOMER-LED DESIGN SYSTEMS

A.O.M. Adeoye*, and T. Szécsi

School of Mechanical & Manufacturing Engineering,
Dublin City University,
Collins Avenue, Dublin 9, Ireland.

ABSTRACT

Much attention has been given to parts classification, coding and retrieval systems. Tremendous work has been done in the retrieval of design images that are similar to the candidate image from the database. However, little has been said or done when it comes to the classification, coding and identification of products that are similar by searching through the database using a wide variety of search parameters, especially of mechanically oriented products in a manufacturing environment. This paper presents a method of classifying, coding and retrieving of identical or similar products from a database for the purposes of Customer-Led Design, based on the principle of Group Technology classification and coding.

The proposed method is based on a hybrid (Chain-Hierarchical type) classification system using a wide range of product parameters. The search and sorting procedure uses two methods for the identification of similar products with each method having procedures for generating product code lists for the similar products, and also applies the Analytic Hierarchy Process (AHP) to analyse and rank the Product Attributes according to their priority vector. The AHP helps in determining the combination weights from subjective judgements regarding the importance of the Product Attributes. The method is successfully implemented in a new Customer-led Design (CLD) environment.

1. INTRODUCTION

It’s a fact of life. Customer demand and technology innovations are increasingly stretching manufacturing capability to the limit [1, 2]. For a firm to be able to compete and survive in this dynamic environment, it must have the ability to rapidly adjust in order to reduce the Lead time needed to design, model, test, manufacture and deliver quality customised product to the customer. In a Mass Customisation environment, the level of customer involvement, and latitude afforded to the customer is necessarily limited, even the products family on offer is limited as well. This is largely driven by the manufacturers’ desire to maintain mass production-type economies and the desire to maintain control [3] and as a result, the flexibility that ought to come along in a co-design setting and customer designing their own products with varieties of products from different family of products has been greatly affected. Consumer are becoming increasingly insistent on having a say in the design of products and looking for an avenue to bring their ideas to reality and coupled with the needed rapid adjustment in the area of manufacturing/product lead time in getting the customised product to them, this has led to this new concept-Customer-Led Design (CLD). And when this concept is applied, the manufacturer tends to save time, cost and labour in re-designing as identical product could be search for from the database and modifies to suit the customer request.

While much have been achieved by manufacturers in the area of customised goods from homes to personal computers, athletic shoes to automobiles, where customers are part or are involved in the designing of the goods or products in order to allow the individualisation of the products with reduction in lead-time and at a reasonable and affordable cost, the same should be extended to products that are mechanically oriented where customers should be actively involved in the designing of their products. Customer-Led Design (CLD) is one way to satisfy this individualisation of the mechanically oriented product. CLD is a customer co-design process of products which meet

* Corresponding author: Tel.: (00353) 8721-84575; Fax: (00353) 1-700-7148; E-mail: adeyinka.adeoye3@mail.dcu.ie
the needs of each individual customer/user with regard to certain product features and all operations are performed within a fixed solution space, characterized by stable but still flexible and responsive processes [4]. Customers are integrated into design creation by defining, configuring, matching, or modifying an individual product - mainly mechanical related products. This is done by the applications of a hybrid (combination of hierarchical and chain-type) classification system to group the products into classes and sub-classes using attribute parameters; the Configuration of products which make it possible to add and/or change functionalities of a core product by suggesting similar products if requested product is not found; a coding system for mechanical designs and applying it to each product according to their position in the hierarchy; and the use of a database to store and retrieve information about the products [5].

The development of Customer-Led Design Systems (CLD) arises from the need to develop effective and efficient systems that will enable the customer/user with little or no technical know-how the flexibility, creativity and influence they desire in bringing their ideal product into reality which were not part of similar systems and also to allow them to select from different family of products in order to build their own product. Established Group Technology (GT) concepts in the area of classification, coding, search and retrieval of similar designs [6] were modified to suit the purpose of the CLD system used. The GT structure has been employed to classify the mechanically oriented products into different applications, and each product was coded using a code system for mechanical design. Having selected a product from various applications in combination with other attributes, the customers build their own individual product and if the initial configuration does not yield any product code from the product database, two methods has been developed [7] to suggest a similar product codes from the product database.

The remainder of this paper is organised as follows. In the second section, an overview of some existing concepts is given that were used to develop this new customer-led design system. Section three explains the details of the CLD system using GT and the test runs for identical product codes. Section four describes the framework and operations of the CLD. In Section five, we conclude and give insight on further work to be done which is based on the present study.

2. OVERVIEW

2.1. CUSTOMISATION OF PRODUCTS

The demand for individualisation of products are on the increase and MC has strives to produce customised products and services on a mass scale with the same levels of responsiveness, quality, efficiency, and costs that are typical of mass production [8] but with little flexibility or constraint that has restricted the involvement of the customer as a co-designer. R. Duray and G.W. Milligan [9] categorised MC based on external involvement of the customer in design, fabrication, assembly or delivery stages of the manufacturing process of the product, all with constraint on the part of customer involvement. In an attempt to satisfy the individualistic taste of the customer based on the product order, a method was adopted to separate products into categories such as Made-to-Stock, Assemble-to-Order, Made-to-Order, and Engineer-to-Order [10] and all of these are example of Customer-Driven “Manipulation”, which is define as the customer/user’s ability to directly influence or manipulate a product’s aesthetic, functional or dimensional characteristics [1].

Although many manufacturing firms, have put MC into practice such as Dell (in computer industry), Adidas Salomon (in footwear industry) [11] and virtual kitchens (in kitchen and cabinet industry) [12], not all the products are mechanically oriented and the degree of customer involvement is limited to pre-defined. In order to fully incorporate the customer into the design process, Z. Siddique and J.A. Ninan [13] designed a web based framework to provide customisable products by integrating the customer into the design process. This would have been a perfect model but due to constraint in the configuration the customer involvement is still restricted and the system does not entail products varieties.

2.2. GROUP TECHNOLOGY (GT)

Group technology was introduced by Frederick Taylor in 1919 as a way to improve productivity and in 1958 Mitrofanov (a Russian engineer) formalized this concept in his book, “The Science Principle of Group Technology”. Ever since, GT has become a very important tool in manufacturing and production systems. This is a manufacturing philosophy that seeks to improve productivity by grouping parts and products with similar characteristics into families and forming production cells with a group of dissimilar machines and processes [14]. The impact of group technology on the production process among other things is: quicker design changes, improved customer service and
building customer relationship through classification. Group technology can be implemented through the following
techniques namely: classification system, coding system and retrieval system. All of these were used for classifying
and coding of product parts which work hand in hand, and retrieving of engineering design parts.

Because of its high effectiveness, the Group Technology technique was adapted for use in the development of the
classification, product coding and retrieval system for this research work in which some modifications were carried
out, in order to be used in the Customer-Led Design system for the consumer products family that are mechanically
oriented.

2.3. PRODUCT DESIGN RETRIEVAL SYSTEM

In the 1950s, a number of systems were developed to facilitate design retrieval. These were systems that allowed
the designer to describe a part in numeric or alphanumeric values. When a new part entered into the design process,
the designer could code it and, with the code designation, retrieve the same or similar drawing from a file cabinet
[15]. Although this system worked in conjunction with classification and coding systems, research has shown that
they are not part of GT applications as was earlier believed [16]. This is because the design retrieval process is
focused on finding the most suitable individual component, not a family of parts as the case may be with GT.

The retrieval system has come of age and as a result, it has found acceptance in the photo-image industry which is
used in retrieving similar product images [17], in manufacturing for retrieving similar engineering drawings from the
database. The wide usage of this system is accounted to the fact that it prevents proliferation of parts or products,
thereby saving cost in designing a new one and also time for creating new parts.

The relevance of similarity retrieval systems to CLD systems is that the author applies the knowledge of the
retrieval system of similar parts from the database, since one of the process of realising the aim of the CLD is by
suggesting a similar product code in the database to the customer/user if the product he/she requested could not be
found through the first interaction.

3. THE CLD SYSTEM APPROACH

The approach adopted for this work is based on the overview conducted as presented above. It can be pointed out
that most, if not all of the research that has been carried out on Customisation is only able to address one issue which
is allowing the customers/users to customized their desired product from a pre-defined template or catalogue, in
which case they are unable to express their idea about their ideal product and thereby constraining their choice.
Though customisation has been implemented in some consumer products, it has not been extended to mechanically
oriented products and services. Secondly, from all indications, Group Technology has been merely based on
classifications and coding of engineering parts and drawings without considering the products in its classification and
coding. The hierarchy used in classification can only work for a particular item, thereby it is limited in its options for
variety of product of family. And finally, the retrieval system as popular as it is, has not been tried or used outside
data, drawings and part retrieval.

3.1. HYBRID SYSTEM OF CLASSIFICATION

The Product Chain System (PCS) is the platform from which the whole concept of CLD emerges. It consist of
product attributes that are developed to form the hybrid (chain-hierarchical) type of classification system which gives
the customer/user the opportunity of a wide-range of product options to select from, and this is based on nine chain-
hierarchy attributes. The GT coding structure consist of Hierarchical, Chain type, and Hybrid codes. The code type
adopted for this system is the hybrid type structure. With the hybrid selected, the authors were able to combine the
mixture of hierarchy and chain type to develop the PCS of the CLD system which makes it comprehensive to permit
for the variety of products which will help the customers in defining exactly what they wanted by systematically
following the defined steps through interaction with the system.

The hybrid system that is used to classify the products is a detailed hierarchical system that makes use of such
attributes as shown in figure 1 below to draw information about the product from the customer (some of the attributes
are explained below). Each of these attributes has classes and subclasses in hierarchical order which helps to narrow
down the customer’s product just like a Decision Tree Classification. The hybrid also allows both vertical and
horizontal cross reference of features or attributes of the products irrespective of other characteristics, that is, relating
one attribute feature to another position of such attribute or sub-attribute in the hierarchy notwithstanding [5]. The
schematic diagram of Product Chain System attributes is shown in figure 1 below.
3.1.1. PRODUCT APPLICATION ATTRIBUTE

This is a function-based attribute that defines the products and classifies them in a hierarchical order according to their uses and is the main attribute of the PCS. The product classes are: Appliance, Sports, Tools and Transportation [18]. However, the application attribute can be enlarged to hold as many attributes as possible. There is no limitation to this function-based attributes. Figure 2 below illustrates the combined weight factor, codes and application attributes.

3.2. MECHANICAL DESIGN CODING SYSTEM

Another part of the CLD is the Coding System for mechanical design which assigns a symbolic description (code) to the products in the Hybrid classification system. The main characteristic that is implored in designing the Products Coding System is flexibility, because customers want to be able to select products from different options and also to build their own products that will suit them and since no two customers have exactly the same needs and that various customers’ needs may have conflicting requirements [19].

In installing the coding system for mechanical design, there are three methods of going about this, by: (1) purchasing a commercial package from a vendor, (2) modifying a publicly available system to fit the required purposes and (3) developing an original system based on the analysis of samples parts [19].

The second method (2) was adapted for the purpose of this work. The Opitz coding system [20] was modified to suit a mechanical design Coding System used for the work. This is because the Opitz coding system is a mixed coding system of Monocode and Polycode in that the monocodes allow a vertical search while the polycodes enable a horizontal, across the board search irrespective of other characteristics [21]. This cross referencing, as was earlier mentioned under the hybrid system of classification, makes the Opitz coding system the suitable choice for the purpose of this work.
The modification that was carried out on the Opitz coding structure is explained below.

### Table 1: Mechanical design coding structure

<table>
<thead>
<tr>
<th>Letter</th>
<th>Attribute</th>
<th>Code Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Applications</td>
<td>Basic</td>
</tr>
<tr>
<td>E</td>
<td>External Motion</td>
<td>Supplementary</td>
</tr>
<tr>
<td>I</td>
<td>Internal Motion</td>
<td>Supplementary</td>
</tr>
<tr>
<td>S</td>
<td>Energy Source</td>
<td>Supplementary</td>
</tr>
<tr>
<td>P</td>
<td>Power Rates</td>
<td>Supplementary</td>
</tr>
<tr>
<td>D</td>
<td>Maximum Dimension</td>
<td>Supplementary</td>
</tr>
<tr>
<td>W</td>
<td>Weight</td>
<td>Supplementary</td>
</tr>
<tr>
<td>F</td>
<td>Shape of Product</td>
<td>Supplementary</td>
</tr>
<tr>
<td>M</td>
<td>Materials</td>
<td>Supplementary</td>
</tr>
</tbody>
</table>

Each item in the Application is given a unique code with a letter A at the front representing the attribute and every other sub-group in the application hierarchy will be represented by a digit according to their position in the hierarchy which might result in a very large number of codes and this will be arranged into a framework such as groups, classes and subclasses, and as a result, customers will be able to select products from various applications to build their own individual product(s). The coding structure for the application attribute is illustrated in figure 2 above. For example, from figure 2 the letter and digits of code A2221 is interpreted below.

Also, the same modality is applicable to other attributes in the Hybrid Classification System with each attribute starting with a letter and then the sub-groups with digits.

To arrive at the Coding Structure used, the following factors were considered; (i) the population of attribute (application, motion, materials, shape, etc.), (ii) the detail the code should represent, (iii) the code structure (chain-hierarchy type), and (iv) the letter representation (A, M, E, etc.) [25].

### 3.3. Methodology for Retrieval of Similar Products from the Database

The objective of the problem at hand is to suggest similar product codes from the product database of the design system, which are similar to the customer request code due to the fact that the initial customer configuration input into the design system is unable to bring out a product code. Two methods have been developed for the above [22].

**3.3.1. First Method**

This method involves two steps. The first step is a search method that uses the first digit of the Application attribute structure of the Product Chain System for the customer/user initial product code to search and retrieve similar product codes from the database. While the second step is a sorting procedure involving a more detailed similarity analysis of the similar products codes identified by step 1. It systematically sorts the similar products codes according to their similarity in terms of weight, shape and materials to the customer/user specified product requested.

**3.3.2. Second Method**

In this method, a Weighted Factor (WF) is assigned to the attributes based on their importance in the PCS. The WF used is summed up to 100% which is distributed to the attributes as shown in table 2 below. This is done using a structural technique called Analytic Hierarchy Process (AHP) [23-25]. This is used to structure and determine the WF value assigned for the ranking of the attributes. Also, each of these values is shared among the sub-groups of each attribute. For instance, Application with 46% shared it among its sub-groups based on hierarchy levels as shown in fig.2 above (the WF value is shown in the bracket in the attribute). The same thing is applicable to other attributes based on the number of hierarchy level they have. With this method, the weighted factor (WF) of the attributes for...
each product is summed up and the database displays the product codes in the user graphic interface ranked according to the WF. The higher the WF value, the more similar the product code is to the customer requested product. And one good thing about this is that the customer/user can indicate the level of similarity he/she wants based on the similarity index.

3.4. APPLICATION FOR RETRIEVAL OF SIMILAR PRODUCTS FROM THE DATABASE

Here, we present the steps followed in order to generate product codes from the customer/user based on the second method which is the most effective and reliable out of the two methods. This is because of its ability to generate product codes that closely similar to the candidate product code.

In generating the product code for a customer/user request and to determine if the code generated is actually what the customer wants, a two step process is used to optimise the model of the product chain system template (fully explained in section 4).

**Step 1: Gather code from users’ specification/configuration.**

In order to gather the product request code from the customer/user, the Product Chain System template as explained in section 4.1. If a feasible solution is achieved in this step, then step 2 is of no importance.

**Step 2: Find closest product code in Hierarchical database:** If the first step does not yield a feasible solution, the design system will try to find similar product codes to that generated initially. This is done through the Product Similarity Retrieval System from the product database by using the parameters inputted by the customer/user when configuring the initial product request.

In order to examine the practicability of the system step, the following test was conducted.

The retrieval attributes shown in table 3 below shows product codes generated for some products stored in the database using the mechanical design coding system. Table 3 is used for the test.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Weight Factor</th>
<th>Distribution Of WF In Hierarchy Sub-Group/Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>46</td>
<td>16 10 7 6 4 3</td>
</tr>
<tr>
<td>Ext. Motion</td>
<td>12</td>
<td>8 4 - - -</td>
</tr>
<tr>
<td>Int. Motion</td>
<td>8</td>
<td>8 - - -</td>
</tr>
<tr>
<td>Energy Source</td>
<td>7</td>
<td>7 - - -</td>
</tr>
<tr>
<td>Power Rate</td>
<td>5</td>
<td>5 - - -</td>
</tr>
<tr>
<td>Max. Dimension</td>
<td>4</td>
<td>3 1 - -</td>
</tr>
<tr>
<td>Weight</td>
<td>5</td>
<td>3 2 - -</td>
</tr>
<tr>
<td>Shape</td>
<td>6</td>
<td>6 - - -</td>
</tr>
<tr>
<td>Materials</td>
<td>7</td>
<td>7 - - -</td>
</tr>
</tbody>
</table>

**TEST:**

Assuming customer/user initial (requested) product code is: **A2221 E22 I2 S1 P0 D21 W11 F2 M5**

This represents a Bicycle, and from the database, there is no feasible product that matches the code, then, by finding the closest product code, we apply step 2 to generate a similar product code to the candidate code.

By applying the second method of the retrieval system, we have the weighted factor as shown in table 2 below.

The product code gathered from the customer configuration data has 100% WF value, i.e. **A2221 (46) E22 (12) I2 (8) S1 (7) P0 (5) D21 (4) W11 (5) F2 (6) M5 (7)**

The Weight Factor value is then generated for the product codes in the database. This is done by ascribing WF values to each attribute of the product codes and the value for each attribute is dependent on their hierarchy level/sub-group as contained in table 2 above. This automatically ranks the product codes according to their WF values with the highest value on top. The higher the WF value, the closer/similar the product code to the candidate code gathered from the customer. Table 3 shows the calculation of WF values for each attribute and the total WF
value for the product in the database, while table 4 illustrates the ranking of the product codes from the highest to the lowest. From table 4 it can be seen that the closer product codes to the candidate one are the codes with WF between 52%– 87% similarity indexes. And this could be suggested as similar product codes to the customer based on his/her configured data.

Table 3: Calculation of the weight factor for each attributes

<table>
<thead>
<tr>
<th>Application</th>
<th>External Motion</th>
<th>Internal Motion</th>
<th>Energy Source</th>
<th>Power Rate</th>
<th>Maximum Dimension</th>
<th>Weight</th>
<th>Shape</th>
<th>Material</th>
<th>Product Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>A211(16)</td>
<td>E1(-)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D21(4)</td>
<td>W11(5)</td>
<td>F5(-)</td>
<td>M5(7)</td>
<td>Exercise Bike(52)</td>
</tr>
<tr>
<td>A31(-)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D21(4)</td>
<td>W11(5)</td>
<td>F5(-)</td>
<td>M1(-)</td>
<td>Wheel Barrow(41)</td>
</tr>
<tr>
<td>A412311(+)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D21(4)</td>
<td>W21(-)</td>
<td>F2(6)</td>
<td>M1(-)</td>
<td>Pallet Truck(42)</td>
</tr>
<tr>
<td>A31(-)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S3(-)</td>
<td>P5(-)</td>
<td>D12(-)</td>
<td>W11(5)</td>
<td>F5(-)</td>
<td>M1(-)</td>
<td>Lawn Mower(25)</td>
</tr>
<tr>
<td>A211(16)</td>
<td>E1(-)</td>
<td>I1(-)</td>
<td>S3(-)</td>
<td>P2(-)</td>
<td>D22(3)</td>
<td>W12(3)</td>
<td>F5(-)</td>
<td>M5(7)</td>
<td>Tread Mill(29)</td>
</tr>
<tr>
<td>A2222(33)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D11(-)</td>
<td>W11(5)</td>
<td>F2(6)</td>
<td>M6(-)</td>
<td>Skate Board(76)</td>
</tr>
<tr>
<td>A2222(33)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D22(3)</td>
<td>W12(3)</td>
<td>F2(6)</td>
<td>M1(-)</td>
<td>Trolley(70)</td>
</tr>
<tr>
<td>A2222(33)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D21(4)</td>
<td>W11(5)</td>
<td>F2(6)</td>
<td>M1(-)</td>
<td>Unicycle(87)</td>
</tr>
<tr>
<td>A412311(+)</td>
<td>E22(12)</td>
<td>I2(8)</td>
<td>S1(7)</td>
<td>P0(5)</td>
<td>D21(4)</td>
<td>W11(5)</td>
<td>F2(6)</td>
<td>M1(-)</td>
<td>Hand Truck(47)</td>
</tr>
</tbody>
</table>

Table 4: Ranking of the database product codes based on the weight factor calculation

<table>
<thead>
<tr>
<th>Application</th>
<th>External Motion</th>
<th>Internal Motion</th>
<th>Energy Source</th>
<th>Power Rate</th>
<th>Maximum Dimension</th>
<th>Weight</th>
<th>Shape</th>
<th>Material</th>
<th>Product Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2222</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D21</td>
<td>W11</td>
<td>F2</td>
<td>M5</td>
<td>Unicycle(87)</td>
</tr>
<tr>
<td>A2222</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D22</td>
<td>W12</td>
<td>F2</td>
<td>M1</td>
<td>Trolley(70)</td>
</tr>
<tr>
<td>A2222</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D11</td>
<td>W11</td>
<td>F2</td>
<td>M6</td>
<td>Skate Board(76)</td>
</tr>
<tr>
<td>A223</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D22</td>
<td>W12</td>
<td>F2</td>
<td>M1</td>
<td>Trolley(70)</td>
</tr>
<tr>
<td>A211</td>
<td>E1</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D21</td>
<td>W11</td>
<td>F5</td>
<td>M5</td>
<td>Exercise Bike(52)</td>
</tr>
<tr>
<td>A412311</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D21</td>
<td>W21</td>
<td>F2</td>
<td>M1</td>
<td>Pallet Truck(42)</td>
</tr>
<tr>
<td>A31</td>
<td>E22</td>
<td>I2</td>
<td>S1</td>
<td>P0</td>
<td>D21</td>
<td>W11</td>
<td>F5</td>
<td>M1</td>
<td>Wheel Barrow(41)</td>
</tr>
<tr>
<td>A211</td>
<td>E1</td>
<td>I2</td>
<td>S3</td>
<td>P2</td>
<td>D22</td>
<td>W12</td>
<td>F5</td>
<td>M5</td>
<td>Tread Mill(29)</td>
</tr>
<tr>
<td>A31</td>
<td>E22</td>
<td>I2</td>
<td>S3</td>
<td>P5</td>
<td>D12</td>
<td>W11</td>
<td>F5</td>
<td>M1</td>
<td>Lawn Mower(25)</td>
</tr>
</tbody>
</table>

4. FRAMEWORK AND WORKING OPERATION OF THE SYSTEM

4.1 PRODUCT CHAIN SYSTEM OF CLD TEMPLATE

The Product Chain System architecture presented below uses the product hierarchy to show the detailed model of the customised products in the database, which will be used by a design engine for analyses, optimisation, and selection of appropriate products/components. To generate product codes and present or display the product from the user selections, a template is used that specifies a set of well defined guidelines and instructions, which guide the customer/user on how to go by and tell the design system in a structured manner how to perform the tasks [14]. The Product Chain System template shown below uses a top down approach to define the family members. On the top of the hierarchy is the Product family, (PF), which comprises of the different family attributes (Pi). Each attribute is defined based on their modules/components and their codes generated by analysing the components/modules in a hierarchical order. The Application attribute (App), being the most important attribute of the product family is analysed using its modules/component in such a way that follows a decision tree type to get the code for the application. The elements of App include – Appliances (Ap), Sports (Spo), Tools (ToI) and Transports (Tra). From these elements, Sport (Spo) module- Indoor (In D) and Outdoor (Out D), is chosen and then analysed to get to the next level. From the analysis, outdoor (Out D) is further analysed to generate wheel (Whl), Sports trolleys (Spo T), and others sport products/equipments (Oth). Also, wheel (Whl) which employed the attribute Option (Whl Op), x and y is analysed. The x option represents wheels that are with drives (WD) while the y option represent wheels without drives (WiD) and the outcome of Whl Op is x. When the App code has been generated, it is then cross-referenced both vertically and horizontally with the other analysed product family attributes to generate the final product codes.
for the customer product. At the lowest level of the hierarchy is the customer-selected product (SP), which represents the end result of the customer specific input to the template.

4.2 SYSTEM OPERATIONS

The basic structure on which the Customer-Led design system will operate will be based on gathering information from the customer/user of the system to generate a customised product code for the customer needed product from the products database. The configuration of the CLD system consists of a customer/user terminal, product database, product similarity retrieval system, and customisation of the product. In this system, the customer/user can bring his/her idea of the type of product he/she wanted to realise through the product chain system of the customer-led design system.

In the basic structure, the customer/user will interface with the product similarity retrieval system and product database by specifying processes on the customer/user terminal. The system application includes the following steps of processing which outlines the system application flow:

Step 1: The user will undergo an interactive section with the system by configuring data about his/her desired product. Here, product information is gathered/generated from the customer/user.

Step 2: The product data inputted by the user into the system is used in generating a product code for his/her desire product. In this step, the product code generated is basically dependent on the customer/user configuration.

Step 3: The Database is then searched for the match of the product code generated

Step 4: If a match is found then, the product code is retrieved from the system to be display on the user interface (GUI) and then if the customer/user still wants to go ahead to customize some of the product features, he/she can proceed. However, if a match is not found in the database, no code will be retrieved and in this case the system will suggest a similar product code using the methodology for the similarity retrieval system explained in section 3.3 above. The methodology will find similar product codes that may have resemblance in application and some other attributes to the one that was requested by the customer/user. Then, step 3 (database is searched) is carried out again and if satisfied, the product is demonstrated on the user terminal. In this manner, the user obtains the desired product from the system.

5.0 CONCLUSION

Using the CLD system for the classification, coding and retrieval of products offers a great market prospective to manufacturers in the universal competitive environment. When manufacturers adapt to this type of design, it will give them an edge in the saturated market of mass and products customisation and also attracts more customers who are more individualistic in their taste because of the flexibility and opportunity to create/design to their satisfaction that are inherent in the design network. It will also save them time, cost and labour in re-designing as identical products could be searched for from the database and modified to suit the customer request.

This paper focused on the application of CLD in classifying, coding and retrieving of products that are similar from the database. This is carried out by the application of GT concepts using a hybrid (combined hierarchical and chain-type) classification system in which the structure is based on coding an application (area of product use) that are functionality based with variety of families and then define the selected application through other products parameters that are inherent in the chain-hierarchy which will eventually lead to the exact or similar group of products that may match the desired product of the customer which are stored in the database.

The project objective has been achieved in that the test carried out in the retrieval of similar product codes using the CLD system was successfully done and customer/user can get as many suggested products codes as he/she wants base on the level of similarity index indicated.

Future work is on going at the moment on how to modify the retrieved product that was similar to the candidate product requested by the customer/user. Also, the modification and designing in terms of design for manufacture and assembly when a product does not appear in the database, developing of the Product/Part Library, and programming of the entire design.
ACKNOWLEDGEMENTS

Dublin City University is a partner of the EU-funded FP6 Innovative production Machines and Systems (I*PROMS) Network of Excellence (www.iproms.org).

REFERENCES

[14]. http://www.freequality.org/sites/www_freequality_org/documents/training/GroupTechnology%5B1%5D.ppt#256,1,Group
[21]. P. Shilpan: Development Of a Classification and Coding System for Computer Aided Planning, Thesis of the Faculty of the Graduate School of the New Jersey Institute of Technology, 1991
Lifetime Models for Condition Monitoring and Simulation Based RAMS Optimizations

Michael Walther1*, Alexander Verl1
Steffen Nebel2, Bernd Bertsche2

1Institute for Control Engineering of Machine Tools and Manufacturing Units (ISW)
Universität Stuttgart
70174 Stuttgart, Germany

2Institute of Machine Components (IMA)
Universität Stuttgart
70569 Stuttgart, Germany

ABSTRACT
Modern production systems are characterised by numerous conjoint and interacting production machines each with sophisticated maintenance and logistic processes. To remain competitive, it is necessary to reduce the system downtimes by applying the optimal maintenance strategy and therefore reducing life cycle costs. Condition diagnostics and condition monitoring often are a very effective way to minimise downtimes of crucial system components, for example ball screws, ball bearings and drives. Since 2006 a research unit at the Universität Stuttgart has been working to develop new approaches, methods and models to improve the predictive maintenance of machine tools, focusing on signals which are already available within the machine control, i.e. without additional sensors as far as possible. One aspect of the ongoing research project is the determination of lifetime defining parameters of ball screws, ball bearings and drives. In addition to the load parameter, the jerk is investigated to improve the existing quantitative lifetime models and to integrate these models into diagnostic procedures to improve the condition monitoring and the prediction of the remaining lifetime of those components. These lifetime models are also taken into account within a superior, Petri net-based reliability model, which allows improving the availability of complex production systems. The model incorporates an extended coloured stochastic Petri net (ECSPN) and a reliability block diagram (RBD) in a conjoint modeling procedure which allows a close-to-reality representation of a complex production system’s behaviour. Besides the modelling of technical aspects, ECSPN allows the analysis of cost aspects such as spare part costs, storage costs and costs of idleness. The analysis of the ECSPN is done by a two-step Monte Carlo simulation method. The modelling tool and the simulation tool have been developed at the Institute of Machine Components.

1. INTRODUCTION
The operational availability of machine tools is an essential prerequisite for the profitability of the manufacturing industry. Maintenance is driven by increasing cost effectiveness. Diagnostics and simulation based optimizations can contribute to a more efficient maintenance of the equipment and to reduced lifecycle costs. It is state-of-the-art to use accessory measurement systems to analyse bearing failure or shaft unbalance conditions. Today, these techniques are applied, often isolated and focused on component or device level. Their returns are not considered in the context of the whole system in order to optimize the operational system availability and costs comprehensively.

At the Universität Stuttgart a research unit, funded by the German Research Foundation, has been established in the year 2006 with the objective of gaining new findings about the subject of predictive maintenance. Members of the research unit are the Institute for Machine Tools (IWF), the Institute for Control Engineering of Machine Tools and Manufacturing Units (ISW) and the Institute of Machine Components (IMA) [1]. Each institute contributes with it’s specific core competence, its models and methods. The four sub-projects deal with experimental wear diagnostics, drive based diagnostics, control based diagnostics and maintenance strategies.

This paper presents a new approach in systems management combining simulation based RAMS (Reliability, Availability, Maintainability and Safety) optimisation of production systems and nowadays condition monitoring methods of machine tools.

* Corresponding author: Tel.: +49 (711) 685-84531; Fax: +49 (711) 685-74531; E-mail: michael.walther@isw.uni-stuttgart.de
2. STATE OF THE ART

The planning and optimization of maintenance activities of technical systems are often done by applying analytical methods which are often limited to independent components, constant failure rates and simple maintenance strategies. The application of simulation techniques often allows a more detailed analysis [2].

Modern production systems are generally characterized by mechatronics, i.e. numerous interacting mechanical and electronic components with software. Their components have various stochastic failure behaviours with constant or time-depending failure rates, for example, the lifetimes of mechanical components can often be described by Weibull or lognormal distributions. The components and devices of such a system are not independent, a failure or overload of a specific component typically influences other components. Common mode failures, like the failure of a superior control system or the loss of power in the whole facility affect all production machines and have to be considered for a close-to-reality analysis of such systems. Also in operational state, production systems offer numerous functional interactions and dependencies, such as load sharing or passive states, which influence the aging of mechanical components over time. Besides the failure behaviour, the maintenance activities of complex technical systems with stochastic delivery delays, limited maintenance personnel, limited spare parts and stocking have to be considered.

In the last decade many condition monitoring techniques have been developed [3][4][5]. They focus on the detection of damages. All sorts of techniques have been used, e.g. state observers, signal analyses and neural networks. These techniques are nowadays well-engineered. Especially those which are based on the analysis of vibration signals, which are picked up by accelerometers, are well-established for rotary machines. In [6] is a definition of how to measure the characteristic parameter describing the wear measuring quantity and its maximum. So far nothing similar exists for machine tools currently.

It is state-of-the-art to detect the condition of local components and devices of a digitally programmed machine tool. The characteristic parameter is used to describe the condition, but the interrelationship between the value and unit of the characteristic parameter and degree of wear is normally not based on a physical phenomenon. The unsolved problem is to predict the remaining lifetime of the production system depending on the characteristic parameter.

A similar problem during the design process of the machine tool was solved by the use of statistical methods and parameters. These parameters depend upon the expected load of the specific component. The methods for calculating the basic dynamic load rating and rating life of rolling components is established and is described in [7][8].

At the moment, these methods for the components with rolling contact do not take into account the jerk of the motion of a CNC (Computerised Numerical Control) feed drive. The load is caused by the force that works on the component. The acceleration of the motion has an influence on the lifetime. Also the sort of friction in the rolling contact is significant for the calculation of the lifetime. So therefore, the time in which the lubricating film is being built up is relevant. It is unknown how the jerk influences the setup of the lubricating film.

LOW-LEVEL PETRI NETS

Within the research unit, a class of high level Petri nets is applied to model technical production systems and their diagnostics.

A Petri net is a graphical modelling formalism for condition-discrete systems which was developed in 1962 by Carl Adam Petri [9]. Petri nets consist of just a few basic elements (essentially conditions and transitions which change conditions) and some simple basic rules which describe the dynamic behaviour of the model. Petri nets can be used wherever dynamic systems with synchronisations, concurrent and parallel characteristics as well as causal dependencies require depiction. The basic elements of Petri nets are;

- Places (drawn as circles, so-called passive nodes within the graph); they represent possible conditions and they can contain tokens.
- Transitions (drawn as squares, so-called active nodes within the graph); they model possible condition transitions and therefore the dynamics within the model following specific rules. After a certain
precondition (the so called activation), a transition can fire, i.e. a transition deletes tokens and creates new tokens in the places which are connected to the transition. If a time delay occurs between the activation and the firing, the transition is called a timed transition, otherwise it is called an immediate transition. Within a stochastic Petri net, the time delay between activation and firing can be stochastically distributed. Therefore, stochastic Petri nets can model a component’s stochastic lifetime or a stochastic maintenance delay.

- Arcs are direct connections between transitions and places and vice versa. They map the logical interrelations.
- Tokens (drawn as filled out circles) are the only dynamic elements within a Petri net. The assignment of tokens of all places are called a markings. A marking describes the current state of the modelled system. The marking therefore changes with a firing transition.

Figure 1 shows a small Petri net model of a simple repairable component with two places representing it`s two possible conditions, ‘operational’ and ‘failed’, and two timed transitions for the switching between these two states, representing the component’s stochastic lifetimes and maintenance times. The figure shows the model in its initial marking in an operational condition. According to the activation rules, the transition ‘component failure’ is being activated and after a stochastic delay it will fire, i.e. the token in the place ‘component operational’ is being deleted and a new token is being generated within the place ‘component failed’, representing the first failure of the component. The new token in the place ‘component failed’ then activates the following transition ‘Component repair’, it’s firing after a stochastic period of repair deletes the token from the input place ‘Component failed’ and creates a new one in the output place ‘component operational’, restoring the initial marking. The component’s lifecycle then starts again. The time slice of the token being in the place ‘component operational’ is equivalent to the operational availability of the component. The analysis of the model is done by applying a Monte Carlo simulation.

Figure 1 – A simple stochastic Petri net model of a repairable component

Since the introduction of the original low-level Petri nets, many new extended classes of Petri nets have been developed to improve their modelling abilities.

**HIGH-LEVEL PETRI NETS**

In 1996, Jensen developed the class of Coloured Petri Nets (CPN) [10]. The innovative aspect of this class of Petri nets is that the previously indistinguishable (anonymous) tokens can carry information and therefore become distinguishable (coloured). The information within the tokens can influence the dynamic of the model by affecting the activation of the transitions. The information within the tokens can be passed on and/or can be modified by the transitions during their firings. Coloured Petri nets allow more compact models, because identical net parts can be reused for tokens with different colours. On the other hand, coloured Petri nets require textual declarations and additional text inscriptions in the graph to depict the flow of information within the model. The depiction of the tokens in a CPN also changes to a text inscription beside it’s corresponding the place.

In 2004/2005, Pozsgai developed a conjoint system model (CSM) consisting of class of Extended Colored Stochastic Petri Nets (ECSPN) and a reliability block diagram (RBD) focusing on the close-to-reality modelling of complex technical systems [11][12]. The class of ECSPN allows modelling of

- arbitrarily stochastically distributed and/or deterministic time characteristics (lifetimes, maintenance durations),
• various system states and component states including passive states, hot/warm/cold standby redundancies and load sharing,
• manifold dependencies between components, the system and the system’s environment, for example common mode failures,
• complex maintenance strategies with corrective and preventive maintenance actions including condition based maintenance actions,
• the failure and maintenance behaviour of pre-aged components,
• queuing behaviour of maintenance/logistic processes and
• activity- and condition-dependent operational costs.

Figure 2 depicts an ECSPN version of the anonymous PN model from Figure 2, consisting of the systems RBD in the upper right corner, the ECSPN component model below and the declarations of the ECSPN in the upper left corner. The token within the model contains the pre-age of the component, which is considered when calculating the component’s stochastic lifetime, i.e. the firing delay of the transition ‘Component failure’, which is depicted by the arc inscription ‘AgeEn = a’. The arc inscription ‘0.15*a’ represents an age reduction of 85% of the repair measurement. The interface between the RBD and the ECSPN is a global component state-variable x1 which is modified by the arc expression ‘Set()’.

3. NEW APPROACHES AND METHODS

3.1 RECORDING COMPONENT LOAD HISTORIES

Based on the online recording data the actual operational wear from every component with rolling contact should be calculated. During the design process of the feed axis the rating life is calculated on a fixed representative motion and process forces. The dimensioning of the mechanical components follow the formulas (1, 2) [8]. For calculating the load (P) for a component with rolling contact (p=3) the force or torque/moment (F) as well as the number of revolutions (n) is needed. For the rating life (L10), the dynamic load rating (C) is given by the manufacture of the component. The time (T) should be set in an adequate and intelligent range.

\[
L_{10} = \left( \frac{C}{P} \right)^{p} \times 10^6 \text{ revolutions.} \tag{1}
\]
This is shown first by the example of a ball screw. The actual mechanical part (illustrated in Figure 3) of a feed drive represents an oscillatory system. Regarded in an abstract way this system consists of discrete masses \( m \), mass moment of inertia \( J \), stiffness \( C \) and damping \( D \) elements. The coupling (lead of ball screw \( h \)) and the value of these mechanical properties is the determining factor for the dynamic behaviour of the feed drive. To describe the dynamic behaviour of the feed drive, differential equations are used. By using the Lagrange's equation for the modelling of the feed drive, the result is the equation of motion. For this to be possible generalised coordinates \( x_1, x_2, x_3, x_4, \phi_1, \phi_2 \) have to be declared. In order to consider damping the Rayleigh dissipation function is added. By using this mathematical method it is possible to describe the linear behaviour of the feed drive.

Figure 3 – Dynamic model structure of a feed drive with ball screw (left side) – Static model for a linear bearing (right side) [8].

To analyse the loads (force, torque/moment) on the component, a static mechanical model has to be calculated. Therefore, it is possible to calculate load rating for every rolling contact. The procedure is shown in [8] by the example of a linear bearing. This is state-of-the-art. In most product catalogues, the component manufacturer describes the usual applications and how to calculate the load. The static model just has to be adapted to the certain feed drive. With the signals that are used in the controller cascade along with the static and dynamic model, the actual component load could be calculated. The position set point of the CNC control unit alone is not meaningful enough. However, because a controller cascade is necessary to get the desired motion, the drive has to put more power into the system according to set point values.

With a state observer using this model, changes of the stiffness and damping during lifetime could be considered. The load is calculated according to formula (2) from the force or moment [8]. By using the modulus of the force and after integrating over the time, the phasing of the force is not relevant. Usually the dynamic behaviour of the mechanics is implicit in the signals of the controller cascade (here, for example, the current). If, for example, a small lead of the ball screw stops the transmission of the dynamic behaviour, a fixed factor should be taken in the formula for the load. Therefore, there is no need to use a state observer for calculating the load. The only thing that is important is the energy that is stored in the rotation motion as well as what is lost in friction. This energy does not
cause a load at the rolling contact. If, in the context of further research an evaluation of the spectral composition of
the load is possible, the developed model or a firmly determined prefactor should be used.

\[ P_{\text{e}} = U \cdot I \]

* If no power is applied, it depends on the mechanics what happen with the energy.

Figure 4 – Use and storage of the power in a feed drive

The size relevant for the load is the amount of the power that is transmitted through the rolling contact. Figure 3
and Figure 4 illustrate the sources, the usage and the transition of the power that is applied to the feed drive. The
allocation of the drive power shown in Figure 4 clarifies this argument. The measurements which are based on
controller cascade signals permit only one sum measurement of all losses caused by friction. Thus the individual
values of the loss moments cannot be determined. The errors caused by ignoring the frictions are not relevant for the
lifetime calculation.

These described methods are implemented in a prototype machine tool. The CNC control unit is used for
recording the load and hosting all relevant information e.g. the dynamic load rating and the load history of every
component [14]. Over an interface, it is possible to access all the information.

3.2. ENHANCED MODELLING

Within the framework of the research unit, the modelling and simulation approach with ECSPN has been
adapted and has been enhanced significantly. Starting from the point of view of a system planner, a number of
system models have been developed and refined iteratively in order to integrate component diagnostic techniques
and models on component and device level. The enhancements allow a close to reality analysis of the complex
maintenance strategies with corrective repairs as well as scheduled preventive and condition based maintenance
actions and dynamic grouping of maintenance activities. Several diagnostic techniques that can be considered are:

- periodic inspections,
- adaptive inspection intervals corresponding to the components aging behaviour,
- dynamically grouped inspection activities,
- monitoring of the current age and age prediction for the time of the following inspection,
- determination of the remaining lifetime and,
• specific detection probabilities and inaccuracy for each condition monitoring system.

When a certain age limit of a component is exceeded or the remaining lifetime falls below a specific limit, preventive maintenance actions can be performed separately or they can be grouped dynamically. In addition, the age of an already failed component can be considered as a decision support within the model. A component with a long operation time might be replaced by a new one, whereas a component which has failed early might be repaired with a certain grade of renewal. By integrating the stocking of spare parts within the system’s model, the influence of the storage capacity and the reorder limits on the system’s availability can be considered adequately.

The key characteristic of the coloured Petri nets (information carrying tokens) allows the integration of analytical lifetime models and degradation models with regard to the load history of the components. Figure 5 shows a lifetime transition of an ECSPN with an input arc. The information of the token is being interpreted as the component’s pre-age and influences the remaining lifetime. When the component changes to another load, or when it is being passivated because of the failure of another component, the actual age is preserved in a token and influences the failure behaviour of the next load level. Figure 6 depicts another lifetime transition in which failure behavior can be varied dynamically during the simulation according to a lifetime model.

The Monte Carlo simulation of ECSPN delivers very manifold and detailed information of the model on the level of the basic modelling elements, namely conditions and condition transitions. The analysis delivers

- inherent component reliabilities and availabilities,
- operational system reliabilities and availabilities,
- the number of failures and maintenance actions,
- operational times and downtimes with their distribution parameters,
- the spare part demand over time and,
- condition and transition based costs and revenues, e.g. spare part costs, cost of idleness, which are added up over the simulation time.

The developed approach offers manifold possibilities for close-to-reality modelling of complex technical production systems, their analysis and optimisations. The method also allows detailed decision support when system configurations, boundary conditions or products change.

5. SUMMARY AND OUTLOOK

In this paper we have presented a new approach for a combined simulation based RAMS analysis of production systems with regard to the current condition of crucial machine tool components. The approach incorporates an extended class of coloured stochastic Petri nets for the modelling of complex technical production systems with an interface to the current component conditions. The analysis of the model is done by applying a Monte Carlo simulation which delivers detailed reliability data, e.g. the operational reliability and availability of components and systems, the spare parts demand, operational revenues and cost of idleness. The approach allows close-to-reality models of complex technical production systems and offers detailed optimisation and decision support possibilities with regard to the system’s current component state. The method can also be applied when the configuration of a system or the product changes. The approach merges the views of the system’s operators/owners, management, reliability engineers and maintenance specialists and delivers profound prognoses.

The next steps include continuing lifetime tests, e.g. to determine the influence of jerk of a motion on the lifetime of ball bearings. The lifetime models will be adapted, extended and refined. The construction of large
complex models will be supported by developing a construction kit of tested model parts with standardised interfaces. The possibilities of modelling and analysing cost aspects will be enhanced significantly so as to consider individual cost calculations.

ACKNOWLEDGEMENTS

The authors would like to thank the Deutsche Forschungsgemeinschaft (German Research Foundation) for the funding of the research group (FOR 639) and Dr. Arbogast Grunau, Member of the Management Board of Schaeffler Group Industrial, for his donation of materials.

REFERENCES

Prediction of Operational Performance for Food Manufacturing Systems Using Discrete Event Simulation and Response Surface Methodology

Ayad K. Mohamed, John Geraghty*, Mohammed S. J. Hashmi,

School of Mechanical and Manufacturing Engineering
Dublin City University
Glassnevin, D9, Ireland

ABSTRACT

The food-processing industry is an important industrial sector. In terms of turnover and employment, it is the largest manufacturing sector in the European Union [1]. The food-processing industry can be considered as a part of the process industries, which is defined as ‘‘firms that add value by mixing, separating, forming or chemical reactions’’ [2]. In food processing, these operations are applied on agricultural raw material to obtain food products. The processing of this raw material can be continuous or in batches [3]. Manufacturing simulation has been one of the primary application areas of simulation technology. It has been widely used to improve and validate the designs of a extensive range of manufacturing systems[4].

The current focus of our research activities is on the prediction of operational performance for food manufacturing systems. In this paper we present an ExtendSim7 simulation model of a simple M/M/n queuing system. The model has been used to study the effect of Inter-arrival Time (range 2 - 5 min), Service Time (range 3–6 min), and number of machines (range 2 – 4 M/c) on the Average Cycle Time, Average WIP, and Average Throughput using Response Surface Methodology (RSM). The experimental plan is based on a Central Composite Design (CCD). Quadratic polynomial equations for predicting the three responses (i.e. the Average Cycle Time, Average WIP, and Average Throughput) were developed. From an ANOVA for the three models, it was found that all the three models are adequate and can be used to predict the responses within the conditions of this experiment. This work has shown that RSM can be used effectively in combination with discrete event simulation as a prediction and optimization technique to further advance our research studies.

1. INTRODUCTION

The food processing industry is an important industrial sector. In 2004, processed food sales worldwide were approximately US$3.2 trillion [5]. In terms of turnover and employment, it is the largest EU manufacturing sector and one of the major pillars of the European economy. In 2005, the food and drink industry turnover in the EU reached €836 billion and employed about 3.8 million workers [1]. In the U.S. consumers spend approximately US$1 trillion annually on food and drink products [2]. The food-processing industry can be considered as a part of the process industries, which is defined as ‘‘firms that add value by mixing, separating, forming or chemical reactions’’. In food processing, these operations are applied on agricultural raw material to obtain food products. The processing of this raw material can be continuous or in batches [3]. Increased demand for better quality products of food, shorter lead times, and reduced costs have encouraged manufacturing organisations to introduce new concepts and changes in production systems to improve efficiency. Simulation modelling and analysis has become a popular technique for analysing the effects of these changes without actual implementation or assignment of resources. Many manufacturing systems can be easily and adequately analysed with discrete event simulation models. However, other systems may require more complex continuous simulation or even combined discrete event and continuous simulation approaches in order to develop valid models [6].
ExtendSim is a powerful, leading edge simulation tool developed and market by Imagine That Inc. By using ExtendSim, it is possible to develop dynamic models of real-life production processes in a wide variety of fields [7]. Response Surface Methodology (RSM) is a set of mathematical and statistical techniques that are useful for modelling and predicting the response of interest affected by several input variables with the aim of optimising this response [8]. RSM is also a useful tool to minimise the numbers of trials required and provide multiple regression approach to achieve both prediction and optimisation [9].

RSM has been widely used by researchers in food sector to predict and optimise the processes, for instance, RSM has been used to optimise formulations of chocolate peanut spread [10], and formulations of chocolate-flavoured, peanut–soy beverage [11]. A new method to control the production rate of a manufacturing system based on the combination of stochastic optimal control theory, discrete event simulation, experimental design, and RSM was presented and implemented in [12]. They succeeded to extend the concept of the hedging point policy which is presented in [13] and [14]. A novel way to integrate the RSM and simulation to develop a worker flexibility policy for a CONWIP controlled serial manufacturing line producing a single class job and operating under saturated demand was presented in [15].

This paper presents a method based on combining ExtendSim7 as a simulation tool and Response Surface Methodology as prediction and optimisation tools to predict the performance of a manufacturing system in order to determine the optimum performance. In this work three input parameters of a theoretical manufacturing system (Inter-arrival time, service time, and number of machines) have been taken into consideration. A design matrix has been developed using RSM. The ExtendSim7 simulation tool was used to build a model of a simple M/M/n queuing system. Then RSM is used to develop mathematical models to predict the performance of the system. The RSM model is validated in comparison to the simulation model.

2. Experimental Work

2.1. Experimental Design

The experiment was designed based on a three level Box–Behnken design with full replication. Inter-arrival Time (range 2-5 min), Service Time (range 3 – 6 min), and number of machines (range 2 – 4 M/Cs) being the system independent input variables. RSM was applied to the experimental data using a statistical software package, namely Design-expert V7. Table 1 shows the design matrix that was developed using the software. Linear and second order polynomials were fitted to the experimental data to obtain the regression equations. The sequential F-test, lack-of-fit test and other adequacy measures were used in selecting the best models. A step-wise and backward regression method was used to fit the second order polynomial equation (1) to the experimental data and to identify the relevant model terms. The statistical software was used to generate the statistical and response plots

\[ Y = b_0 + \sum b_i X_i + \sum b_{ij} X_i^2 + \sum b_{ij} X_i X_j \]  

(1)

2.2 Development of the Simulation Model

ExtendSim7 has been used to build the simulation model. The simulation model is a discrete event simulation, so the ‘Item library’ within ExtendSim, which is designed for building discrete event models of commercial and industrial processes, has been used as the primary source for modelling blocks. Figure 1 below shows a screen-shot of the “M/M/n” model developed in ExtendSim. 17 experiments were conducted on the simulation models, each experimental condition was replicated ten times.
Prediction of operational performance for food manufacturing system

Table 1: Experimental Design Matrix

<table>
<thead>
<tr>
<th>Exp. no</th>
<th>A: Inter-arrival time</th>
<th>B: Service-time</th>
<th>C: no. of m/c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4.5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>4.5</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>4.5</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>4.5</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>3.5</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

2.3 Development of the Simulation Model

ExtendSim7 has been used to build the simulation model. The simulation model is a discrete event simulation, so the ‘item library’ within ExtendSim, which is designed for building discrete event models of commercial and industrial processes, has been used as the primary source for modelling blocks. Figure 1 below shows a screen-shot of the “M/M/n” model developed in ExtendSim. 17 experiments were conducted on the simulation models, each experimental condition was replicated ten times.

![Figure 1: Snapshot of ExtendSim 7 Model Developed for Experimentation](image)

Table 2: Experimental Measured Responses

<table>
<thead>
<tr>
<th>Exp. no</th>
<th>Avg-C.T</th>
<th>Avg-WIP</th>
<th>Avg. Throughput</th>
<th>Exp. no</th>
<th>Avg-C.T</th>
<th>Avg-WIP</th>
<th>Avg. Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.042</td>
<td>1.037</td>
<td>340.977</td>
<td>10</td>
<td>111.263</td>
<td>23.932</td>
<td>218.087</td>
</tr>
<tr>
<td>2</td>
<td>3.004</td>
<td>0.512</td>
<td>170.581</td>
<td>11</td>
<td>3.304</td>
<td>0.734</td>
<td>222.018</td>
</tr>
<tr>
<td>3</td>
<td>7.493</td>
<td>2.481</td>
<td>331.687</td>
<td>12</td>
<td>6.115</td>
<td>1.403</td>
<td>228.566</td>
</tr>
<tr>
<td>4</td>
<td>6.064</td>
<td>1.028</td>
<td>169.892</td>
<td>13</td>
<td>4.357</td>
<td>0.962</td>
<td>220.967</td>
</tr>
<tr>
<td>5</td>
<td>209.742</td>
<td>60.522</td>
<td>290.345</td>
<td>14</td>
<td>4.462</td>
<td>0.975</td>
<td>217.474</td>
</tr>
<tr>
<td>6</td>
<td>13.582</td>
<td>2.862</td>
<td>210.578</td>
<td>15</td>
<td>4.476</td>
<td>0.993</td>
<td>221.538</td>
</tr>
<tr>
<td>7</td>
<td>4.584</td>
<td>1.528</td>
<td>333.429</td>
<td>16</td>
<td>4.422</td>
<td>1.004</td>
<td>227.141</td>
</tr>
<tr>
<td>8</td>
<td>4.367</td>
<td>0.720</td>
<td>164.565</td>
<td>17</td>
<td>4.476</td>
<td>0.993</td>
<td>221.538</td>
</tr>
<tr>
<td>9</td>
<td>6.644</td>
<td>1.903</td>
<td>284.775</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. RESULT AND DISCUSSION

In this research, three responses were measured: Average Cycle Time, Average WIP, and Average Throughput (see Table 2). This section discusses the results of the experiments. In section 3.1 the Analysis of Variance results are presented and discussed. Section 3.2 discusses the validation of the metamodel and finally section 3.3 discusses the effect of the process parameters on the responses.

3.1. ANALYSIS OF VARIANCE

The test for significance of the regression models, the test for significance on individual model coefficients and the lack-of-fit test were performed using the statistical package Design-Expert V7. By selecting the step-wise regression method, which eliminates the insignificant model terms automatically, the resulting ANOVA (e.g. see Table 3) for the reduced quadratic and 2 Factor Interaction (2FI) models summarising the analysis of variance of each response and showing the significant model terms can be obtained. The same tables show also the other adequacy measures R² which is a measure of the amount of variation around the mean explained by the model, adjusted R² which is a measure of the amount of variation around the mean explained by the model, adjusted for the number of terms in the model. The adjusted R² decreases as the number of terms in the model increases if those additional terms don’t add value to the model, and predicted R² which is a measure of the amount of variation in new data explained by the model. The predicted R-squared and the adjusted R-squared should be within 0.20 of each other. For all three response variables, all the adequacy measures were close to 1 and the difference between the predicted R² and adjusted R² is less than 0.2, indicating the models are adequate fits.

The adequate precision compares the range of the predicted value at the design points to the average prediction error. In all cases the values of adequate precision are dramatically greater than 4. An adequate precision ratio above 4 indicates adequate model discrimination. The analysis of variance indicates that for the Avg. Cycle Time model, the main effects of the Inter-arrival Time (A), Service-Time (B), and Number of Machines (C), and the second order effect of Service-Time (B²) are the most significant model terms associated with Avg. Cycle Time., the interaction of A and C, and the interaction of B and C are considered to support hierarchy. In this model the lack-of-fit is not significant at “Prob > F” =0.002. Also, in the case of the Average WIP model the main effect of the Inter-arrival Time (A), Service-Time (B), Number of Machines (C), the second order effect of Service-Time (B²), the interaction of A and B, the interaction of A and C, and the interaction of B and C, are the most significant model terms associated with Average. WIP. Finally, for the Average Throughput 2FI model, the analysis indicates that the effect of the Inter-arrival Time (A), Service-Time (B), Number of Machines (C), the interaction of A and C, and the interaction of B and C are the most significant model terms associated with Avg. Throughput.

Table 3: ANOVA table for Avg. WIP reduced quadratic model

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F Value</th>
<th>p-value Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3.8913</td>
<td>10</td>
<td>0.3891</td>
<td>475.9469</td>
<td>&lt; 0.0001 significant</td>
</tr>
<tr>
<td>A-Inter-arrival time</td>
<td>0.8622</td>
<td>1</td>
<td>0.8622</td>
<td>1054.6100</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B-service-time</td>
<td>0.8951</td>
<td>1</td>
<td>0.8951</td>
<td>1094.7584</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>C-no. of M/cs</td>
<td>1.9858</td>
<td>2</td>
<td>0.9929</td>
<td>1214.4399</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>0.0442</td>
<td>1</td>
<td>0.0442</td>
<td>54.0904</td>
<td>0.0003</td>
</tr>
<tr>
<td>AC</td>
<td>0.0706</td>
<td>2</td>
<td>0.0353</td>
<td>43.1584</td>
<td>0.0003</td>
</tr>
<tr>
<td>BC</td>
<td>0.0279</td>
<td>2</td>
<td>0.0139</td>
<td>17.0491</td>
<td>0.0034</td>
</tr>
<tr>
<td>B²</td>
<td>0.0147</td>
<td>1</td>
<td>0.0147</td>
<td>17.9247</td>
<td>0.0055</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0049</td>
<td>6</td>
<td>0.0008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit</td>
<td>0.0036</td>
<td>2</td>
<td>0.0018</td>
<td>5.4280</td>
<td>0.0725 not significant</td>
</tr>
<tr>
<td>Pure Error</td>
<td>0.0013</td>
<td>4</td>
<td>0.0003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cor Total</td>
<td>3.8962</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.999</td>
<td></td>
<td>Pred R-Squared</td>
<td>0.974</td>
<td></td>
</tr>
<tr>
<td>Adj R-Squared</td>
<td>0.997</td>
<td></td>
<td>Adeq Precision</td>
<td>84.044</td>
<td></td>
</tr>
</tbody>
</table>
3.2. Validation of Model

The final mathematical models in terms of coded factor as determined by design expert software are as below:

\[ \frac{1.0}{\text{Avg CT}} = 0.16 + 0.016 \times A - 0.077 \times B - 0.11 \times C[1] + 0.059 \times C[2] + 0.018 \times AC[1] - 7.354 \times 10^{-3} \times AC[2] + 6.281 \times 10^{-3} \times BC[1] - 0.014 \times BC[2] + 0.019 \times B^2 \]  

(2)

\[ \frac{1.0}{\text{Avg WIP}} = 0.74 + 0.31 \times A - 0.32 \times B - 0.54 \times C[1] + 0.14 \times AC[1] + 0.082 \times AC[2] + 0.076 \times BC[1] - 0.068 \times BC[2] + 0.059 \times B^2 \]  

(3)


(4)

Note: Suppose we want to predict the three responses while the category factor at its lower level (i.e. two machines), replace C[1] with 1 and C[2] with zero and verse versa for three machines, however, for four machines replace both C[1] and C[2] with -1.

Figure. 2, Figure. 3, and Figure.4 illustrate the relationships between the actual and predicted values of Avg. C.T, Avg. WIP, Avg. Throughput, respectively. These figures indicate that the developed models are adequate because the residuals in prediction of each response are small, since the residuals tend to be close to the diagonal line.
3.3. **Effect of Process Parameters on the Responses**

### 3.3.1. Average Cycle Time

It is evident from Figure 5, which is a perturbation plot in a transformed scale (inverse) showing the effect of each factor for two machines, that the Inter-arrival Time (A) has a negative effect on the Average Cycle Time. On the other hand, the Service Time (B) has a strong positive effect on the response. The same trend has been observed for three machines with a significant reduction in Average Cycle Time. In the case of four machines the same effects have been observed with no significant reduction in Average C.T. This means that there is no significant difference between systems containing three and four machines, in parallel, within the range of our input parameters.

### 3.3.2. Average WIP

In the case of two machines the analysis indicates that the Inter-arrival Time (A) has a strong negative effect on the Average WIP, where the service-time (B) has a positive effect. The same trend has been observed for three machines and four machines with less severity. Figure 6 shows the effect of factors on the Average WIP for three machines. Again the analysis shows that there is no significant difference between systems containing three and four machines, in parallel, within the range of our input parameters. Hence, this is an important point as by using less number of machines and still getting the same output, this behaviour would result in reducing the total cost without significant effect on the overall process performance measures.

### 3.3.3. Average Throughput

The analysis indicates that for the system with two machines, both factors A, and B have a strong negative effect on the response. While in the case with three machines factor A (Inter-arrival time) has a significant negative effect on the response, and factor B (service time) has a very slight negative effect on the response, as shown in Figure 7. For four machines the same behaviour of factor A has been observed and factor B has a slight positive effect on the response, this indicates that the second level of factor C (number of machines) might be the optimum level.
4. CONCLUSION

This paper presents a simple application of Discrete Event Simulation and Response Surface Methodology for the analysis of simple $M/M/n$ queuing systems in which the effects of Inter-arrival times of jobs, service times of machines and number of identical parallel machines on average cycle time, average WIP and average throughput have been investigated. The paper shows that RSM can be used effectively as a prediction and optimisation tool in food manufacturing system, and the Box–Behnken design can be employed to develop mathematical models for predicting in performance of the manufacturing system in general or in food manufacturing system. The prediction analysis indicates that the second level of factor C (number of machines) might be the optimum level of this factor. Additionally Factor A (Inter-arrival Time) at the second level of Factor C has a negative effect on all responses. Finally, Factor B (Service Time) has a positive effect on the Avg. Cycle Time and Avg. WIP.

This paper represents a first step in our research programme. It verifies the usefulness of combining RSM and Discrete Event Simulation to develop metamodels for manufacturing systems analysis. The next stage of the research programme is to develop simulation based metamodels of 'long-lines' operating under various production control strategies such as CONWIP, Kanban and Hybrid CONWIP-Kanban and verify their applicability to food processing industries. The models will be used to investigate the optimisation of such systems under uncertainty of the demand profile, service reliability, and component/raw material availability and in the presence of bottleneck operations.

5. ACKNOWLEDGEMENTS

Libyan Government is gratefully acknowledged for the financial support of this research.
6. REFERENCES

[1] CIAA, Data & Trends of the EU Food and Drink Industry, CIAA (Confederation of the Food and Drink Industries of the EU) 2005.


Analysis of capability of Lean Manufacturing implementation
- case study

Anna Sobiś, Joanna Olesków-Szląpka, Marek Fertsch
Poznan University of Technology in Poland
Institute of Management Engineering

ABSTRAKT
Manufacturing companies have done big step forward within last years - change is visible in range of useless reduction of productive ability and within the confines of logistical chain. These operations lead correction of operative efficiency and simplification of flows of value streams, they also assure fast reaction of clients demand. However, is it possible to employ Lean Manufacturing (LM in text used further) for value streams that featuring of few clients and big disparity of final products?

The purpose of this article is an analysis of capability of Lean Manufacturing implementation in manufacturing enterprise from sector of electromechanical industry for which characteristic is big variety of final good and complexity of productive process. Authors will indicate in article main problems, which have met while implementing LM and will discuss basic tools that find application in this particular enterprise.

1. INTRODUCTION
Object of analysis is manufacturing enterprise from sector of electromechanical industry (as a WAC in text described further) that implements elements of Lean Manufacturing. It deals with producing industrial parts, which process flow is based mainly on mechanical processing - also chemical or thermal treatment. Several processes of special types exists in given firm: soldering, sand-blast cleaning of surface, honing. Unfortunately, no manufacturing line is autonomous, that mean that many cooperation are taken in productive process, internal equal as well as external. It effects negatively on organization of work and production managing in this company. To reinforce market position (in particular to fir to increasing clients’ demand), the company has decided to implement certain of LM elements. Question will appear here surely “Why only certain of LM elements?” - answer will be placed in farthest elaboration part.

2. SIPOC ANALYSIS
The most important in initial phase of operation is to understand problems on the shop floor and familiarize with a value stream. Value stream is “specific set of actions that are realized in supply chain and wanted for designing, ordering and delivering specific product or service” [2]. Therefore, there are all necessary actions for process raw material to finished product. It is possible to use for definition of flow of value instrument SIPOC, which is taken advantage in methodology Sigma Six. SIPOC diagram is usually applied with reference to processes, however it is possible to use it for values streams. SIPOC supplies visual information about process/value stream – it tells how it is realized to supply value clients [8]. Acronym SIPOC comes from English and it means:

- “Suppliers” that provide input in your process
- “Input” to define the material, service or information that is used by the process to produce the outputs
- “Process” that your team is improving; typically it is a defined sequence of activities that will add value to inputs to produce outputs for the customer
- “Outputs” are considered the products, services, and information the is valuable to the customer.
- “Customers” that use the outputs that are produced by the entire process” [8]

When creating a SIPOC diagram, your project team does not necessarily need begin at the beginning. In fact, the team should probably start in the process (P) phase and ask questions about the process itself. The team should label the process with the summaries of the most critical three to six steps. After analyzing the process, they should
document what (O) is delivered to whom (C). The team can brainstorm and prioritize the most critical one to three customers. They then can identify, prioritize and align the outputs most significant to those customers. The next step would be to verify these initial assumptions with voice of the customer tools from the DMAIC process and/or designate as critical to quality, speed or cost. Lastly, the team can identify what input or information (I) is needed to perform that process and who provides that input (S). This brainstorming and prioritization of significant inputs finishes the activities around building a SIPOC.

In summary, there are three main reasons we many times begin the Six Sigma process by building a SIPOC diagram. These are [3]: a SIPOC diagram quickly and easily captures the current or "as is" state of the organization and processes in question; the SIPOC exercise brings associates together in a non-threatening way that builds; the SIPOC exercise allows the team to review all the processes in a way that they can easily see which next steps can be identified. Steps to complete the SIPOC diagram are very easy to complete. Here are the steps you should follow [3]:

- Create an area that will allow the team to post additions to the SIPOC diagram. This could be a transparency (to be projected by an overhead) made of the provided template, flip charts with headings (S-I-P-O-C) written on each, or headings written on post-it notes posted to a wall.
- Begin with the Process. Map it in four to five high level steps.
- Identify the Outputs of this Process.
- Identify the Customers that will receive the Outputs of this Process.
- Identify the Inputs required for the Process to function properly.
- Identify the Suppliers of the Inputs that are required by the Process.
- Optional: Identify the preliminary requirements of the Customers. This will be verified during a later step of the Six Sigma measurement phase.
- Discuss with Project Sponsor, Champion, and other involved stakeholders for verification.

A typical SIPOC diagram is a 5-column tabular format. If possible, have your project team start at the beginning of the process phase. The team should ask numerous questions in the process phase about the actual process itself. The SIPOC diagram will capture the as is state of your organization and the processes in question. The SIPOC diagram will allow your team to review all the processes in a visual manner so they can easily identify which steps need to be identified [8].

By using SIPOC diagram for value stream analysis in given manufacturing enterprise from sector of electromechanical industry image presented below (Figure 1) was received:

![Figure 1. SIPOC Diagram – Manufacturing Value Stream](image)

Source: own research based on [4]
From SIPOC diagram follows that Q company is supplier of raw material and X company is final recipient – therefore here exists one external supplier and one external client. Value stream, that is presented on the foregoing figure, consist of several (generally expressed) periods: ordering raw material, its supply (and taking in the firm), technical preparation of production, according to earlier processed technology production of products and finished products dispatch. Inputs and outputs from value stream are described and assigned to the specific periods in value stream.

3. CLIENTS’ DEMAND ANALYSIS FOR WAC COMPANY

As already have been mentioned, one external client exists in considered case, who makes demand to SAP – orders for supplier of raw material are built on its base. According to the Figure 2 clients’ demand can be steady (A), seasonal (B), high-variation (C) or low-variation (D) [1].

![Figure 2. Client’s demand](source: [1])

Above-mentioned diagrams were compared with clients’ demand in WAC company. On Figure 2 there are shown clients’ demand for whole WAC company - small decrease of order of clients is visible in 2008. From diagram presented on Figure 3 follows that demand in given company is almost constant – however, it features not big changeability – it could be compared to example A from Figure 2.

![Figure 3. Client’s demand for whole Company W](source: own research)

However if we compare clients’ demand between 2007 and 2008 for particular manufacturing line (SV line in text used further) there can be noticed greatest divergence. In 2007 clients’ demand was kept middling at the level of
568 pieces per month while in 2008 it totaled about 503 pieces per month. Small decrease of clients’ order was observed in one year. For SV line clients’ demand is changeable and it is featured low variability – it could be compared to D example from Figure 2.

![Client’s demand: 2007-2008](image)

**Figure 4. Client’s demand for specific manufacturing line (SV) in WAC company**

Source: own research

In whole scale of manufacturing plant, the firm has incorporated orders for about 1025 different productive parts in 2008 and 1092 parts in 2007. All of products is characterized dissimilar process flow (less or more compound) – families of production parts were characterized to ordered them. Identification of product family relies on: setup of similar process flows to groups, configurations and material, so that setup of line dedicated was advantageous and proper.

It finds mirror in amount of different products fabricated between 2007 and 2008 for SV line (Figure 5). In 2007 there existed from 20 to 37 active products, however in 2008 – from 13 to 36. Middling, on SV line there is 27 productive parts started up each month.

![Comparison of quantity of active products in 2007-2008](image)

**Figure 5. Quantity of active products in 2007-2008 (SV line)**

Source: own research

4. **P-Q Analysis**

According to the literature [1], it is possible to define manufacturing company by analyzing amount of fabricated (different) products and amount of final clients – (Figure 6).
The most easy to implement LM principles is a manufacturing company with low number of clients and producing undifferentiated products (example 1 on Figure 6) or company with many clients but still producing almost the same goods (example 4 on Figure 6). However, applying this type of change in companies that produce very different products is really very hard – like in given WAC firm (it could be compared with the example 2 on Figure 6). Typical example for case 3 from Figure 6 is Tool Room, which usually has a lot of clients (external and internal) and simultaneously there are no serial production neither recurrent products – every time device that is produced is fitted for individual requirements of clients. Implementation of LM principles in such a background is almost impossible.

Quoted in article WAC enterprise is a image of firm with small amount of final client and big disparity of produced products. It is very difficult to apply LM in such a company. There is a big scale of production, including a fact that WAC company does not have at disposal autonomous manufacturing lines, it often caused problems at realization of LM principles.

First barrier is lack of autonomous lines. Here appears necessity of cooperating with other productive lines or with external firms –in case of special processes mainly. It means also necessity of withdrawing in a process on the same machines, that were used in earliest phases of processes. Fitting layout of manufacturing line in “U” letter would not bring awaited effects. WAC company has low budget and cannot allow to purchase additional machines. Besides, the product track would not be significantly improved (cut) – in this case the best solution is prepared P-Q analysis (product-quantity) and use spaghetti chart and design layout on this basis. P-Q analysis is “one of the simplest tool of real analysis of volume of production or storage, it is a point of exit for accommodation project design of equipment on the shop floor and warehouse. This juxtaposition should be ordered from biggest amount for smallest, then carried out Pareto analysis (called ABC analysis)” [5]. Such analysis for particular manufacturing line was performed – result is shown on Figure 7.
P-Q analysis should be applied with process matrix and could be used to determine different types of products – A group (20% parts builds 80% quantity), B and C. Sometimes companies decide to create a separated productive line for parts with the biggest volume (A group) – however, it is not possible in WAC firm. Therefore appears an attempt of fitting productive line for current requirements.

5. SOLUTIONS AND PROBLEMS IN WAC COMPANY

Tools, which were implemented to improve material flow, are mainly: TPM, 5S, SMED, VSM, POKA-YOKE (they are described more widely in [7]). It was started with 5S to well organize each work station and discovered problems that were hidden under the dust. After that it was time for VSM (Value Stream Mapping) sessions on each manufacturing line to find out waste and highline the major problems with material and information flow. On the basis of information from the shop floor there was created a current state map - KAIZEN bursts were used for pointing problems and located them in value stream. Those data were necessary to prepare action plan and future state map. In this plan were many issues connected with machines – i.e. if there was to long SETUP, the best tool to solve this problem was SMED; if there was a setback with machine uptime there was used the TPM to prevent machines’ breakdowns. POKA-YOKE (improving product quality by preventing defects) was implemented to eliminate human errors on work stations. Usually there were implemented low-cost devices that were developed right on the shop floor by production workers. Not always it was level 1 (prevent errors), sometimes it was level 2 (detect defects) or even level 3 (manage questionable product made). But what is worth mentioned is the fact that those tools allowed to better material flow on manufacturing lines. Amount of unplanned downtimes has fallen off after arrangement of permanently TPM review and time that Maintenance staff devoted on repair of entered failure were shorten. At the beginning time for repair was about 7 hours (in 2004), it felt off successively in next last years: 4.5 hours in 2005, 3.5 hours in 2006, and it has been maintained at the level of 2.5 hours in 2007 and 2008. It indicates on positive trend in operations from range of preventive service of machines. Besides, in WAC company are often organized VSM sessions, which helps to locate waste in each manufacturing line and prepare action plan. It is frequently refer to applying LM tool like: POKA-YOKE or SMED. Reducing setup time (SMED) brings reduction of non-value added time - it means decreasing waste. Similar in case of application of solution from range of human errors prevention (POKA-YOKE), which allow to cut down amount of defect products and increase safety of executable works. 5S techniques have mainly contributed to creation of friendly work-place. Except of correction of shop floor visualization, they have effected positively on working posts - i.e. using tray shadows. These actions have contributed to improvement of work organization in every day. However, the company has difficulties to maintain continuous flow by using OPF (One Piece Flow) or through FIFO queues (First In First Out). Time analysis for Takt Time, setup time and producing time was done for each manufacturing line - in majority case setup and producing time were not balance with Takt Time. OPF is implemented when setup time and producing time for one piece of product is equal or very approximated to Takt Time. Then exist a possibility to develop employees’ cross-functionality so joined operations could be executed by specific staff. There should be taken under consideration fact that there is no autonomous manufacturing lines, that additionally disturb flow of material. The biggest difficulty causes, in case of FIFO queue, imperative principle that after filling FIFO queue there cannot be produced next part of material on previous machine. Certainly it is tied with limitation of WIP (Work In Process) and simultaneously it clash with employees’ firm goals. Efficiency of employee is calculated from amount of elaborated baseline hour (metric depends on quantity of produced products) - so, they care deeply about achieving the best result. Not producing means that they do not get bonus purposes, it is equal with resignation of additional financial benefit - more widely this problem is described in [6].

6. SUMMARY

In real life production managing is tough and brings a lot of difficulties. Even if there is a great theoretical knowledge the experience and shrewdness is needed to be up to harsh reality. Especially now, when is hard situation on the market, we can observe that there is no golden mean to bring companies out from the crisis. Therefore any tool that could restrain waste on the shop floor is desirable. In this particular company were used only some tools like: TPM, 5S, SMED, VSM, POKA-YOKE. They help to reduce non-value added activities in value stream but as was mentioned before those tools did not create flow. The WAC company should be focus on material flow - so far the main task is reducing WIP (Work In Process) and producing according to the Takt Time.
REFERENCES

[3] http://www.businessknowledgesource.com, What is a SIPOC diagram (High level process map) and how it is used in Six Sigma, z dn. 13.02.2009
Discrete event simulation for manufacturing system analysis: an industrial case study

A. Gatsou, X. V. Gogouvitis, G.-C. Vosniakos*

School of Mechanical Engineering, Manufacturing Division, National Technical University of Athens, 157 80, Greece

C. Wagenknecht and J.-C. Aurich

Institute for Manufacturing Engineering and Production Management, University of Kaiserslautern, Germany

ABSTRACT

A real-world manufacturing system is studied in this paper, producing fitness training machines, which are typically used in gyms. Several variants of these machines are produced to stock, but overall lead time was considered to be too long. Therefore, although the manufacturing system structure was considered stabled, it was studied using discrete event simulation software in order to validate initial explanations and also in order to try new manufacturing control policies, i.e. to tune the system. The manufacturing system consists of a saw, two turning station, a machining centre, a drilling machine, welding stations and a powder painting station. It was decided to follow and draw up the processes associated with one particular variant, because production separates between variants anyway, and also because it would be impractical to construct a simulation with all possible variants in sequence. For the variant studied all processing, setup and waiting times were recorded by monitoring real production over a period of two months. An as-is simulation model was constructed referring to batch size of 200, equal to the order size of this product variant. The main problems that were found in the system were: first that scheduling was conducted by production supervisors intuitively without concrete criteria, second that waiting times were excessive in input and output queues which were physically concentrated in a central storage area, and third that most machine operators were under-utilised. Therefore, the model constructed was amended to provide improvement alternatives. For instance, process transfer batch sizes were cut by 75%, processing priorities were assigned according to the respective processing times and one more operator was added in the drilling station. In this way lead times were cut by 1/3, waiting times were reduced from 6.8 to 3.5 days on the average etc. The particular problems associated with the realistic nature of the study are outlined.

1. INTRODUCTION

In organizing a middle-sized company, the adequacy and efficiency in production resources – raw material, machinery, capital, plant and human resources – play a decisive role. Strategies such as Total Productive Management (TPM) [1], the Toyota Production System (TPS) [2] and Just-in-time methodology (JIT) [2] have been identified and implemented as the best practices of production optimization. These approaches of lean thinking base their success in effective management of all available resources in order to create value for the customers.

To materialize these different methodologies of process amelioration a variety of simulation programs are used, simplifying the decision making process for the engineer. These simulation programs vary; they can be used for production process mapping, an output analysis and production predictions. This new market trend is referenced in many researches as digital factory or virtual production and constitutes the main ingredient of scenario simulation [3]. In the USA, the Delmia virtual reality simulation program [4] is vastly used amongst several defense companies and NASA. On the same note, in an attempt to simplify production, many engineers propose to break the given system into three categories [5] which will be separately simulated: product, procedure and resource. Overall, there

* Corresponding author: Tel.: (0030) 210-7721457; E-mail: vosniak@central.ntua.gr
is a general effort to gather all the different simulation milestones in a virtual library [6] in order to have a point of reference for future studies.

The objective of this research focuses on the production of a middle sized manufacturing company in Pirmasens, Germany. Ergo-fit GmbH produces exercise machines for gyms and hospitals. The main goals were:

- to identify the different manufacturing processes and the material / information flow between them
- to record the bottlenecks that occurred in the production of a single exercise machine (namely, Cross 3000)
- to simulate the current state of the production process with the use of simulation program
- to extract conclusions from the simulation’s output analysis
- to develop different scenarios based on the proposals made and simulate them.

2. METHODOLOGY

Shanon, Gordon and Law [7], [8], [9], proposed a methodology regarding the necessary steps included in a simulation-based analysis. As can be seen in Figure 1, the researcher needs to start out by planning the study. In Phase I, the analysis of the manufacturing system, using widely accepted tools (e.g. Gantt, PERT, Petri nets), is conducted. This phase includes the definition of the system and the building of a model representing the system. The importance of these steps is undisputed; false analysis of the real manufacturing system will result in a misleading simulation process. Thus, it is suggested that the researcher gathers all necessary data, identifies the relation between causes and effects, detects the influential factors, and distinguishes the different procedures and their constraints. In addition, in order to achieve essential focus, the usage of correct data and sources and any logical assumptions, where applicable, to validate the accuracy of the data is highly stressed. Phase II deals with a realistic representation of the manufacturing system defined in Phase I. The researcher is then in position to run the simulation, to analyze the results and to record the conclusions. In any step of the process, the researcher is allowed to return to previous steps and proceed to amendments; the methodology proposed is repetitive. This empowers the researcher to analyze different scenarios of the manufacturing system, fine tuning aspects of it and optimizing the system according to different variables.

![Figure 1. Simulation-based analysis methodology](image)

In our case, in order to simulate the given production situation, necessary data needed to be collected and all the production processes needed to be mapped out in detail. Thus the daily observation of the production of the specific product (Cross 3000) was very informative and provided us with the necessary data such as cycle times of the different machining centers, stock in between processes, elapse times, number of scraps per process and more. The data was recorded in a daily log along with other observations such as delivery status of the raw material, avoidable time or material wastes, changes in production scheduling, stock status in different stages, etc. Different data acquisition techniques were used in order to have a clear and complete picture of all the company’s procedures. The company’s standardization protocol (ISO) was also studied, assisting us with more data, while checking if it reflects reality. Furthermore, we focused on the Production Department with detailed analysis of its programme, the dead-times and processing times, the transferring of semi-ready parts of the products, as well as the administration of available resources. The results of this analysis were compared to model tables that were provided from the Production Planning Department.
This set of information is the basis of the simulation model used in order to receive an output analysis of the current production state. Throughout this procedure one of our main sources of information were the people who work in the production processes involved. With the broader information gathered in the beginning of our research for all of the departments of Ergo-fit, the vital facts given to us by the employees as well as the measurements collected, we were able to have a clear and detailed understanding of the manufacturing processes involved in the development of the product.

The next step after designing the system and defining the elements that compose it was to provision the simulation program with the measurements collected during our three month production observation. Promodel [10] is an easy-to-use simulation tool with direct results and visual representations of these results. Its friendly environment allows the user to define all the elements of production such as the machining centers used, number of resources, paths used in transporting the semi-ready parts, batch sizes, and even prioritize the sequence of these events (Figure 2).

After all the measurements have been put into the system and the model has been built, the user may move onto simulating it. If the simulation cannot be conducted due to programming errors, Promodel has a very helpful function that identifies false programming and proposes alternative solutions. The simulation of the current production status functioned as expected and the results gathered in the form of logs and graphs helped us identify the different bottlenecks in the production procedure. After carefully examining the reasons for these production delays we produced several scenarios and tried to implement them to Ergo-fit’s production line by simulating and comparing the results.

Promodel is an ideal tool to simulate and analyze a production line, experiment with different scenarios in the manufacturing procedures and plan a whole production from scratch. However the results concluding each simulation are quantitative and it is in the user’s best interest to interpret these accordingly and come up with suggestions that will have a qualitative impact on the production processes at hand.

![Figure 2. Promodel’s layout module](image)

### 3. Manufacturing System Analysis

Figure 3 shows the flow of material and information inside the company. Customers go through the sales department to place an order. The department then communicates with both the production and assembly department to see if the order can be fulfilled immediately or when the product(s) can be available. In the second case, the production department is informed about the new orders and proceeds in contacting the production planning department to gather information regarding raw material. The production planning department checks if there is enough raw material in stock; if not, they proceed to ordering raw material. They also amend their –at the time– six-week production plan to include the new order.

The manufacturing system of the production departments consists of a number of processes, depending on the part. Initially, all raw material passes from the saw; these parts are then transferred to small storage areas near the next process (drilling, milling, turning), which differ from part to part, or the main storage area (Storage 300). Parts need to end up in assembly, passing through any of the intermediate stations (Storage 300, welding, powdering). Once the parts are assembled, the final products are transferred to the assembly storage. Sales department is then informed that the order is in place and the customer is notified to pick up the products.
Following the steps outlined in §2, we proceeded in a three-month daily observation of the production line and kept critical data on a log. Specifically, we recorded all time related data for every part (waiting time, setup time, process time) and compared them to the ProAlpha times provided by the production planning department. Based on our observations we were able to pinpoint the system’s deficits. For example, there was clearly enormous waste of time during ordering raw material due to vague understanding of the storage’s status; checking of the storage is done manually and there is no update unless requested by the production planning department. Most of the processes present adequate set-up and process times, while there are large waiting times in Storage 300 due to lack of daily production programming. Furthermore, there is no definite understanding of the sequence of processes for every part, while prioritization is handled by the production manager on the spot and with inexplicit criteria. Storage 300, while being one of the most important areas in the system, is not territorially defined; it basically covers most of the shop floor. This situation is problematic, since it is highly difficult to locate parts which were placed there a long time ago, while forklifts have great difficulty in transporting parts. Welding is frequently stalled, due to waiting of all the necessary parts to arrive, although process time is better (faster) than ProAlpha provided time.

Additionally, since the assembly department is working primarily based on the programme supplied by the sales department, while the production department is working based on the programme supplied by the production planning department, issues regarding material flow arise. If there is not enough space for parts to be stored in the assembly storage (e.g. low sales mean more parts waiting in the storage), the production department needs to slow down, whereas if data shows that storage will be empty before previous estimates, the production department needs to speed up to cover the deficit. Clearly, this problem can be solved only by reorganizing information flow between the different departments.

4. MANUFACTURING SYSTEM SIMULATION

4.1. MODELLING THE CURRENT MANUFACTURING SYSTEM (“IST”)

After monitoring all the company’s activities, especially those regarding the production line, we were in position to set up a simulation model. First, we define the system’s locations, which are the machines that are used and the storage areas (queues). Then, we define the entities of the system, namely the raw material that enters the system, the semi-ready parts, and the final parts. Entities do not follow the same routes, e.g. some parts go through welding whereas others do not. Raw material enters the system by defining the arrivals module of the software, based on the ordering table, which provides the quantity and the time they have to enter the system.
The next step involves defining the resources (namely machinists) and the path networks. The distance among the locations and the speed of the movement among them does not play an important role at this phase, when compared to the waiting times. Finally, we define the system’s processing, which is the production logic followed, while taking into consideration the following remarks:

- Waiting time in queues before the processes
- Setup time of the locations
- Processing time of every part
- Moving time (if available, compared to waiting time).

Counters are also used in critical places inside the system, e.g. before or after a location. They provide us with interesting data while simulating regarding the quantity of entities that are inside the system or processed by a machine, assisting in diagnosing a mistake in processing steps. Finally, we define the shifts and breaks of the resources. Every shift is 7 hours and in most locations there are two shifts.

After simulating the model, we are provided with a variety of reports regarding the simulation’s results. First, a general report is generated, including data for all the variables (locations, entities, resources, counters). These data regard utilization percentages for all the variables, as well as quantity data for the entities. We are also provided with information for every variable separately, via the usage of diagrams. Some of these diagrams were used to acquire interesting quantitative results, namely Location utilization, Multiple Capacity Location States, Multiple Capacity Location Downtimes, Single Capacity Location States, Resource Blocked in Travel, Resource Utilization, Resource States, Entity States etc. Another option is time based results, either via diagrams or histograms with bars, monitoring an entity’s phase changes. Overall, simulation results provided a quantitative analysis of our quality remarks.

As far as locations are concerned, it is safe to state that all machines do not have significant down times. Their maintenance is done on a regular basis off production times and problems arise seldom. Comparing their diagrams with ProAlpha data, we conclude that the saw and the drill are processing faster than expected, while processing times of the lathe or the mill are slower and their set-up times do not comport with the times provided by the production department. Regarding waiting queues and storage locations, the biggest delays are before entering the saw (max: 13 days, average: 6.3 days) and after the saw (max: 13 days, average: 4.25 days).

As far as resources are concerned, although Promodel cannot provide data regarding the workers’ productivity, since it calculates only the utilization of a resource while assigned to a machine, we can still compare its data with ProAlpha numbers and come up to certain conclusions. The worker assigned to the saw has little time working on the saw, since its process is fully automated. This worker is also responsible for the whole Production Department, hence he has no specific working place. The worker assigned to the drill shows better performance compared to ProAlpha data, while having the largest moving times of entities from and to the drill. The worker assigned to CNC machines (either the lathe or the mill) spends more time than expected to set up the machine, while remaining almost 50% of the total time idle, waiting for the machine to finish the process.

The entities state graph presents the results for all the entities, comparing them with each other. Keeping in mind that a number of entities may be describing the same item, providing us with information regarding the different phases that the item is going through, we can see exactly where the item is stalled (e.g. waiting). Also, since entities are also grouped into new entities when moved in batches from one location to another, it is more than obvious that group entities are only in a moving state. The rest of the entities are either waiting or in process. The entities that are about to enter the drill have the biggest waiting values, due to lack of resources, whereas there is significant waiting before CNC processing. Waiting times for raw material entities and for entities before the saw are rather low.

Finally, general reports of the simulation are used to ensure that all the entities have exited the system, meaning that there are no entities in queues or in processing in any location after the simulation has ended. Another way to validate our simulation is the comparison between simulation time and time that was recorded from monitoring real production. The total time, starting when raw material entered the system until the exit of the final product of our simulation is almost equal to the time of the real production line. Differences exist due to variant break times of the resources, changes in the resources’ production rate, and overtimes that were not recorded during our observation of the real production line.

Overall, using our observations and the simulation results, certain remarks were made to improve the productivity of the company. Those remarks can be summarized as follows:
The company’s production system is not adequately defined. The assembly department assembles parts according to customer demand, thus the procedure can be described as a “pull system”, picking up parts from the production department’s stock. At the same time, the production department is producing according to directions from production planning department, following a “push system”, in the sense that it is pushing finished parts into the storage before the assembly department. The size of every batch is either 200 or 400 parts, producing 200 final products. Cutting those batches to half or more would decrease waiting times and storage space, which would benefit the company, from a financial point of view.

There are large waiting times in multiple capacity locations (storage, entry/exit queues, etc.) between processes. This is due to lack of programming and inadequate processing logic that is used for the processing steps of every part. Establishing a more efficient process plan would minimize those waiting times.

Every day observation showed that there is no specific storage area. We propose one large area where raw material will be placed, plus three (3) smaller storage areas close to the processing machines (saw, drill, CNC machines).

Most resources have rather large idle times, since most of the processes are automated. Those resources should be otherwise engaged during processes, e.g. in quality control, setting up another machine, etc.

4.2 Improved Production Model (“SOLL”)

Taking into consideration the analysis of §3 and the results of our first simulation (§4.1), we proceeded in making changes to our simulation model, in order to valuate improvement in productivity. The arrivals of the entities in the system were in groups, according to the product they are related to. The order of arrivals was determined according to the time the entities need to “stay” in the system; in other words, the more time they are in the system, the earlier they arrive. Another important change is that we have all the locations in the layout, with different setup time for every process, while machine capacity is 1 and queue capacity is infinite. There are five machinists used (resources); four of them are used explicitly to each machine, whereas Machinist #1 is not bound to one machine (the saw, in this case), since he is Head of Production Department at the same time. Changes in path networks refer to adding more routes, so that all locations are connected to each other. As far as the processing table is concerned, the most important change was with regard to the waiting time of entities in various queues; those times were minimized, since we expect every entity to be moved as soon as possible. All other time-related activities (setup, processing) were modelled using the command WAIT. We also use variables / counters in order to be able to monitor as many as possible activities and entity changes, while paying more attention to the counter monitoring the arrival of raw material (entities enter the system) and the counter monitoring the entities that exit the system. Figure 4 shows how every entity is processed and the paths they take. Entities can be seen only when simulation is running, since they are dynamically altered.

![Figure 4. Complete layout of scenario “soll”](image-url)
Once the simulation was conducted, we used the same kind of diagrams to evaluate how the changes affected certain aspects of the system. As far as locations are concerned, we can see improvement in all locations but the queue before the drill. The input queue of the saw has huge improvement, while data for the output queue are not available, since entities are transferred immediately after processing. There is significant raise in waiting times before the drill, which is something expected; the process times of the drill are relatively larger than the other processing times, therefore entities coming from the saw have to wait for the drill to become available, leading to large waiting times. As far as the input queue of the CNC is concerned, although the average waiting time is 1.5 days, something that occurs due to large processing times, the improvement is huge, when compared to the initial 5.3 days.

Moreover, one has to be extremely careful when comparing the results of the two simulations. In the first simulation, we gathered data from simulating one model per product part or family, whereas in the second simulation the data refer to the complete production system. In that sense, waiting times in the second case are more valuable, since an entity might enter the system very early but will still have to wait for previous entities to be processed, leading to higher waiting times.

As far as resources are concerned, most resources are used satisfactory, although the machinist occupied in the saw has again very low utilization due to the aforementioned reasons.

The results regarding the entities present an interesting picture. Starting off, general reports assure us that all the entities have exited the system, all queues are empty and there are no entities in any location at the time the simulation ends. Since there are no waiting times between processes, some entities tend to be blocked in queues. The maximum blocking time was 10% of the total time of an entity in the system, whereas the average time was 1%. Comparing entities to each other via the state utilization diagram, we observe that the largest waiting time is before the drill; this waiting time is also characterized as “waiting for available resource”, meaning that the resource of the drill (Machinist 3) is processing another entity. That result, combined with the fact that the company has one more drill that is not being used, allow us to proceed to further improvement. Thus, we amend our model with one more drill and one more resource to utilize it and we proceed to simulation of the model.

Simulation of the third model yielded lower waiting times for the drills’ input, as expected, whereas the total utilization of their resources equals the utilization of the one resource in the previous model. Still, we amended our model once more, fine tuning the process transfer batch sizes, reducing them by 75%. These changes were made under the assumption that high waiting times occurred due to parts waiting for other parts to finish processing before batch transfer. As can be seen in Tables 1 & 2, simulation of the fourth model validated this assumption. Waiting times were reduced more, while there were also changes to resource utilization (smaller batch sizes equals more transfers, thus more resource utilization).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Location</th>
<th>Saw_in</th>
<th>Saw_out</th>
<th>Drill_in</th>
<th>CNC_in</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Min</td>
<td>Avg</td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1</td>
<td>6.3</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0.87</td>
<td>0</td>
<td>0.24</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Locations’ waiting times (max/min/average) for all scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Resources</th>
<th>Saw</th>
<th>Lathe</th>
<th>Drill (1)</th>
<th>Drill (2)</th>
<th>Mill</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.17%</td>
<td>21.77%</td>
<td>51.31%</td>
<td>-</td>
<td>29.84%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.17%</td>
<td>21.77%</td>
<td>34.27%</td>
<td>17.04%</td>
<td>29.84%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.10%</td>
<td>18.14%</td>
<td>32.40%</td>
<td>28.36%</td>
<td>29.57%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Resources’ utilization for scenarios 2 – 4
5. Concluding Remarks

In this paper, a real-world manufacturing system was analyzed and simulated. Following the proposed methodology, we were in position to detect the system’s weaknesses both through shop-floor observations and by using simulation tools. Furthermore, we outlined certain amendments to improve the system’s efficiency and productivity. Assigning accurate process scheduling and defining precise and smaller storage areas were some steps that improved results under study. Redefining information flow among departments to ensure that production is conducted according to the aforementioned schedule is also a proposed step. Simulating the amended models showed that we achieved significant decrease in all waiting times and faster production of the parts. Realizing these amendments in the real manufacturing system should lead to more efficient, ultimately just-in-time production, supporting the production strategy of the company.

Acknowledgements

The work reported in this paper is partly funded from the 03ED217 research project, implemented within the framework of the “Reinforcement Programme of Human Research Manpower” (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Fund).

6. References

On the Suitability of the Information Fusion JDL-U Model as a Reference Model for Virtual Product and Production Development

Leo J De Vin*

Centre for Intelligent Automation & Virtual Systems Research Centre
University of Skövde
Box 408 541 28 Skövde, Sweden

ABSTRACT

This paper presents a description of Modelling & Simulation as used in the Virtual Systems Research Centre and issues discussed in previous FAIM and other papers, such as phases in a simulation project, Verification, Validation & Accreditation, and the use of simulation as a tool to reduce uncertainty. The role of the human in various phases/activities in simulation projects is highlighted. Two models for Information Fusion, the JDL and JDL-U model, are discussed. Subsequently, the activities and phases in a Modelling & Simulation project are placed in the context of the JDL-U model. This comparison shows that there are very strong similarities between the six (0-5) levels in the JDL-U model and activities/phases in Modelling & Simulation projects. These similarities lead to the conclusion that the JDL-U model with its associated science base can serve as a novel reference model for Modelling & Simulation.

1. INTRODUCTION

This paper builds on previous FAIM papers such as “Information Fusion for Simulation Based Decision Support in Manufacturing” [1], “Verification, Validation and Accreditation for Manufacturing Simulation” [2], and “The Role of Simulation in Engineering Design and Production Development Processes” [3]. The aim of the current paper is to put virtual product and production development and the various phases that can be distinguished in a simulation project in the context of the Information Fusion JDL-U model [4] which is an extension of the original JDL model [5]. As such, the paper should be regarded as a discussion paper proposing a novel theoretical framework that links Virtual Manufacturing and Information Fusion much more tightly than the FAIM2005 paper [1] does. Another difference is that the FAIM2005 paper links Virtual Manufacturing to the OODA loop (also known as Boyd’s loop) and not to the JDL model as is done in the current paper. Whilst the OODA loop is a more adequate model to describe Information Fusion (IF) in general, the JDL model is more suitable as a theoretical reference model for various phases in a simulation project. JDL-U model is an extension of the original JDL model and takes into account the user aspects of IF systems. This is presented as an additional level called “User Refinement”. This extension is related to the area of “Human Competence World” in virtual product and production development as discussed in the FAIM2007 paper [3]. A summary of relevant parts of these previous FAIM papers and some other papers/documents will be given below in Section 2.

2. A SUMMARY OF IMPORTANT SIMULATION ISSUES AND THEORY

2.1. A DEFINITION OF “SIMULATION”

There are many definitions of "simulation", but within the context of this paper, the author would like to describe simulation as "experimentation with some model of a system of interest (SoI)". This SoI may be an existing system, a projected system, or a completely imaginary system. Simulation helps us to reduce uncertainty in the decision making process by enabling us to explore many alternatives against relatively low costs and by presenting knowledge about projected artefacts on various levels of abstraction. Simulation as a communication tool also

* Corresponding author: Tel.: +46 500 448000; Fax: +46 500 448599; E-mail: leo.devin@his.se
Flexible Automation and Intelligent Manufacturing, FAIM2009, Teesside, UK

enables us to obtain a refined specification of customer requirements by exposing prospective customers to alternative design solutions in a virtual environment. A prominent question in any simulation is how much we "trust" the simulation results. This role of simulation in the innovative industrial process also makes us more aware of the lack of a suitable science base for engineering work. At best, people try to apply modified theory of science to engineering. This, however, results in an obvious mismatch as science deals with asking questions about the existing whereas engineering is about creating the (presently) non-existing [6].

In a working paper of the Virtual Systems Research Centre at the University of Skövde, five main different uses of simulation and virtual systems are distinguished, of which the first two and the fourth are the most relevant to the area of virtual product- and production development:

1. Simulation as a tool to study an SoI in order to create new knowledge about the modelled system, or to refine existing knowledge about it. This new or improved knowledge can subsequently be used for decision support.

2. Simulation as a tool to train operators in the use of the SoI. In this case, the simulation model serves as a means to transfer knowledge about the SoI to the operators.

3. Serious gaming as a way to create situations that are realistic, even although the situation itself may never occur. This can be used to train people and organisations for situations in which communication, coordination and decision making are important, such as in complicated rescue actions or natural disasters.

4. Simulating a situation and/or sequence of events in order to test peoples’ attitudes or responses. This includes responses of people to (prototypes of) new products.

5. Gaming primarily as entertainment. Nevertheless, such games can include various educational moments.

2.2. PHASES IN A SIMULATION PROJECT

A number of activities can be distinguished in a simulation project. Most authors present these activities in the form of a flow chart. However, such flowcharts should not be treated too rigidly, as a lot of zigzagging between activities/phases may take place. An example of such a flowchart is given in Figure 1 [7]. Short descriptions of the activities/phases are also given in for instance [8].

Fig 1: Activities in a simulation project (after [9] and similar decompositions, for instance [10])

2.3. VERIFICATION, VALIDATION AND ACCREDITATION/Acceptance

REVVA [11, 12] is a methodology for Verification, Validation & Accreditation (VV&A) of simulation models. This methodology is summarised and adapted to manufacturing simulation in [2]. Although most simulation terminology is common for many (sometimes very different) simulation domains, a brief glossary can be useful:

- Modeling & Simulation (M&S). Either means the process of building a model and carrying out a simulation, or denotes the result of this process. In the latter case, the M&S is seen as a product.
- System of Interest (SoI). The existing or imagined real-life system that one wants to study with the M&S.
On the Suitability of the Information Fusion JDL-U Model as a Reference Model for Virtual Product and Production Development

- Verification. The process of demonstrating that the model has the property “correct” (one has built the model right).
- Validation. The process of demonstrating that the model has the property “valid”, usually through comparing a system’s known output with simulation output under similar input conditions (one has built the right model).
- Accreditation/acceptance. The decision of certifying that a model is correct and valid, or the decision that there is sufficient evidence about a model being correct and valid to assume that it is correct and valid.

The relationship between the System of Interest and its associated model is shown in Figure 2. The model is an abstract representation of the SoI and ideally, from the behavior of the model, conclusions can be drawn concerning the SoI. Likewise, from the observed behavior of the SoI, conclusions can be drawn concerning the validity of the model by comparison and interpretation.

2.4. Simulation as a Tool to Reduce Uncertainty

In [3, 13, 14] it is explained why accurate knowledge about non-existing products is not possible. Furthermore, behavioral indistinguishability between a model and a SoI cannot be demonstrated beyond doubt. Hence when we simulate, we use a model that may be incorrect to study a SoI that may have other properties than the ones we envisage or expect. However, simulation can be used to reduce uncertainty regarding the SoIs behaviour. The simulation may not generate knowledge about the envisaged SoI, but it still provides information of the type “if the model is correct and valid, then the SoI will have properties (p, q) under conditions (x, y, z)”. Hence, the role of simulation is not to obtain knowledge about the envisaged SoI (as this is principally impossible), but to provide information regarding the (possible) consequences of certain design solutions. As such, it is a tool for informed decision making which helps the decision maker to analyze the potential effects of alternative decisions and to reduce the level of uncertainty in the decision making process. Figure 3 illustrates these relationships between Real World, Model, and Decision. Fagerström [15] puts the use of models of an SoI in a concurrent engineering context which inevitably implies the use of models; for instance one cannot study the impact of the introduction of a new (to be developed) product variant in a production facility without using models of these. However, the nature of the models can vary; they can be mental models, mathematical models, 3D models, and so on.

3. The Importance of the Human in Simulation Projects

Humans can play different roles in simulation projects. They can for instance be “subject matter specialist”, “modelling & simulation expert” or “contextual user”. Humans in the first two roles play an important role in
building simulation models and those in the third role typically are decision makers. In order to explain the importance of the first two, below we present an expanded and refined version of Figure 2.

![Figure 4: Refined view on Model and SoI](image)

A major difference between this representation and the REVVA representation from Figure 2 is that it is impossible to compare model results to actual system behaviour. This is due to principle limitations of human perception [13, 16]. As a result from this, the model results can only be compared to perceived system behaviour. This results in perceived validity indicators for the model. However, even this comparison is subjective in itself. A typical scenario would be that M&S experts perceive a model as valid, whereas subject matter specialists perceive a model as flawed in some aspects. A possible remedy to this potential conflict of interests is to appoint independent/external VV&A executioners [11, 12]. For the contextual user, the most important validity indicator is whether the model results are an adequate basis for informed decision making. Since different types of decision require different information, to the contextual user (and thus, to other M&S stakeholders as well) validity of a simulation model and its results always have to be related to a specific purpose. Even when the contextual user is not an M&S expert, the user must have sufficient competence to order a simulation project or model.

Apart from issues related to VV&A, M&S experts and subject matter specialists also have to address building a model of the right level of detail. Especially relatively inexperienced simulation modelers have a tendency to build models that are too large, detailed and unnecessarily complex. This is partly due to the fact that this (building large and detailed models) is seen as a merit, and partly it is caused by a desire to satisfy subject matter specialists. Even if the subject matter specialists may not always say so explicitly, it is often thought that detailed models that accurately reflect reality will increase their tendency to accept the model as valid. However, detailed models take longer to build, they generally require more, and also more detailed data, and for some applications such as simulation-based optimization for on-line decision support, execution of detailed models may take too much time. They may also provide too many options to the contextual user and present simulation results with too much detail. As a result from this, the contextual user may find the simulation too complicated to run (too many options have to be set) or may find the results not clear enough for informed decision making. An example of a simulation tool for decision support that takes these issues into account is the FACTS Analyzer [17] presented at the FAIM 2008 Workshop.

4. **THE JDL AND JDL-U MODELS FOR INFORMATION FUSION**

Information Fusion (IF) encompasses the theory, techniques, and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensors, databases, information gathered by human, etc.) such that the resulting decision or action is in some sense better (qualitatively or quantitatively, in terms of accuracy, robustness, etc.) than would be possible if these sources were used individually without such synergy exploitation [18]. An example of IF in manufacturing is the fusion of information from multiple sensors [19]. Examples of IF in the context of this paper can be the fusion of information from the past operation of a manufacturing system (e.g., stored in databases), from the present (e.g., sensor signals, machine status), and from the future (in particular, predictions obtained through simulations). High level IF often has the purpose to provide decision support to a human decision maker (or group of decision makers). In this respect, it has similarities with modeling and simulation for decision support, as will be explored further on in this paper.
A model initially developed for data fusion (an important sub-area of information fusion) is the so-called JDL model. In 1991, the Joint Directors of Laboratories in the U.S., with input from the information fusion research community’s leaders, developed a data fusion paradigm. This paradigm, shown in Figure 5, aimed at providing a framework for communication and coordination amongst the many diverse fusion workers. The initial names of the levels are based on the initial military defense context, however a more general terminology is proposed in [20]:

- **Level One Fusion Processing – Object Refinement.** Level one processing combines parametric data from multiple sources to determine the state and other attributes or identity of low level entities.

- **Level Two Fusion Processing – Situation Refinement.** Level two processing develops a description or interpretation of the current relationships among objects and events in the context of the operational environment. The result of this processing is a determination or refinement of the operational situations.

- **Level Three Fusion Processing – Strategic Refinement.** Level three processing develops an extra-organizational oriented perspective of the data to estimate extra-organizational capabilities, identify opportunities, estimate extra-organizational intent, and determine levels of risk.

- **Level Four Fusion Processing – Process Refinement.** Level four processing monitors and evaluates the ongoing fusion process to refine the process itself, and guides the acquisition of data to achieve optimal results. This includes interactions among the data function levels and with external systems or the operator to accomplish their purpose.

Level 2 is sometimes also referred to as “situation awareness” and Level 3 as “impact analysis”. Usually, a Level 0 Processing is also distinguished [5]. This Level 0 Processing deals (in military terms) with sub-object data. It incorporates for instance data analysis, resolving data conflicts, and conversion of data (for instance, to a common format or to common time & location frames). Llinas [5] elects not to incorporate a Level 5 Processing (User Refinement) as this is not yet an established extension within the Information Fusion research community. However, within the context of this paper, Level 5 is highly relevant and will be discussed briefly below.

Level 5 Processing as proposed by Blasch and Plano [4] in their JDL-U (with “U” standing for “user”) model includes determination of who has access to information or queries information, and presentation of information to support cognitive decision making and subsequent actions. Important factors in Level 5 processing are workload, attention (including guiding and controlling attention), and trust in automated systems. Humans typically generate a mental model that describes purpose and explanations of system functions from observed states, and predicts future states. Recurrent situational scripts and scenarios are experiences that will contribute to the development of the human from novice to expert.

![Figure 5: The initial JDL Model](image)
5. THE JDL-U MODEL AS A THEORETICAL FRAMEWORK FOR SIMULATION PROJECTS

5.1. COMPARISON BETWEEN JDL-U LEVELS AND SIMULATION PROJECT ACTIVITIES/TASKS

With the purpose of high level IF and M&S often being similar, namely providing support for informed decision making, it would appear to be obvious to compare IF and M&S. In essence, in both cases, there is a human in the loop decision process. However, even on the lower levels, the similarities are striking. Figure 6 shows how various activities in a simulation project relate to the six (0-5) levels of the JDL-U model. Since the six levels of IF have already been discussed before, below we name the level and then describe the corresponding level/activity in an M&S project.

![Figure 6: Comparison of JDL-U levels and M&S phases](image)

**Level 0 – Source Pre-Processing**

The corresponding activity in an M&S project is very similar to that in an IF process, such as gathering and analyzing data, resolving conflicts, and removing multiple entries of the same data. Low level data mining can also be executed here to unravel patterns i.e. low level relationships between data. Data of less significance can be discarded.

**Level 1 – Object Refinement**

The corresponding activity in M&S is model verification, i.e. an analysis of whether the building blocks of a simulation model are correctly implemented. This can be compared to correct identification of objects (in the original military IF terminology).

**Level 2 – Situation Awareness**

The corresponding activity in M&S is model validation, i.e. an analysis of whether the model as a whole behaves in a way that is trustworthy. In essence, this means that the behaviour of the model is compared to the behaviour of the real system under controlled conditions, compared to theoretical behaviour (e.g., using trends and lower/upper bound analysis), or compared to results from a previously validated model (e.g., a model with a higher level of detail).

**Level 3 – Impact Analysis**

At this level, predictions about future states and their impact are made. The similarities between IF and M&S are most striking for discrete event simulation projects, in which different production layouts, different planning solutions, or the effects of the introduction of new products or product variants can be studied. However, more generally speaking, this level deals with prediction of future states and thus also accommodates for instance analysis of the impact of design solution on product properties/features.

**Level 4 – Process Refinement**

Just like for IF, the first four (0-3) levels typically deal with single projects or scenarios, whereas Levels 4 and 5 mainly pertain to improvements made from one project/scenario to another. For M&S projects, this means building models of the right level of detail, increased insight into which data are crucial and which are less relevant, speeding up model building through modularization, and so on. Process refinement in M&S is thus fed by comparing simulation results with actual outcomes of implemented solutions, case studies in which alternative models over the same system are built, and so on. This level mainly pertains to M&S experts and to a somewhat lesser extent to
subject matter specialists. However, contextual user aspects such as defining the right problem and ordering the corresponding simulation study can also be seen as part of process refinement.

**Level 5 – User Refinement**

Whereas Level 4 mainly pertains to M&S experts and subject matter specialists, Level 5 mainly pertains to both these and to the contextual user. An example of user refinement is improved distribution of tasks and responsibilities (although this can also be done within a single simulation project). Another form of user refinement is the creation of trust in simulation projects amongst the contextual users. In order to achieve this, simulation projects must yield results that to some extent are predictable, in the sense that similar studies should yield similar results. Apart from this consistency, the results should also form an adequate and dependable basis for informed decision making.

**5.2. Examples of Level 1 and Level 2 Entities**

Examples of entities on Level 2 can be production cells or lines in the case of discrete event simulation. This is in principle the level that is subject of the simulation study. The corresponding entities on Level 1, i.e. one level lower, are for instance machines or stations. In the case of computer aided robotics (geometry simulation), machines or stations are Level 2 entities whereas devices are Level 1 entities. In analogy with terminology from ISO 15 228, these can be seen as Systems of Interest and System Elements of Interest respectively. In principle, this can be a recurrent decomposition as shown in the line-machine-device example above. A corresponding example from product development could be a MBS (multi-body simulation) model incorporating condensed FEM (finite element modeling) models of flexible components which in turn have elements as sub-system entities. The Level 2 entities are also the entities that are of most interest for Levels 3-5.

**5.3. Comparison and Refinement on Levels 3 through 5**

Just like “the proof of the pudding is in the eating”, the proof of simulation results is in how well they compare to observed behaviour of implemented solutions. In cases where simplified models were used, their suitability to offer adequate decision support can be judged by a comparison with results obtained from more detailed models – provided that these more detailed models have been validated previously. A critical assessment of simulation results can result in Level 4 Process Refinement, such as determination of required level of detail of models, data, required number of iterations, or length of the transient state in case of simulation that need a “warm up” time. In the longer run, these assessments result in the development of human skills, competences and professional attitudes across various stakeholders. Professional attitudes include objectivity during validation phases, and correct use of simulation studies. For instance, using a simulation study to provide decision support, instead of using it for retrospectively motivating an already taken decision [7].

![Figure 7: Comparison between simulated and observed results on Level 3 contribute to Level 4 and Level 5 refinement](image)

**6. Concluding Remarks and Outlook**

High level Information Fusion and Modelling & Simulation for product and production development have in common that they provide decision support for human decision makers. One prominent Information Fusion model is the JDL model together with its extended version, the JDL-U model. The comparison between the JDL-U model and the activities and phases in simulation projects has revealed very strong similarities on all six levels distinguished in the JDL-U model. From these similarities, it can be concluded that the JDL-U model can serve as a novel reference model, or at least a basis for such a novel reference model, for Modelling & Simulation. As a result from this, a logical conclusion is that it will be worthwhile to explore similarities between Modelling & Simulation and the JDL-Ú model further in the future, in particular on the higher levels such as Process Refinement and User
Refinement. Nevertheless, the similarities on the lower levels are also striking. Thus, it is fair to expect that large portions of scientific theory associated with the JDL-U model, such as attention models and skills & competence models will also be applicable to Modelling & Simulation projects. This would mean that the science base that supports Modelling & Simulation projects will increase significantly.

REFERENCES


Work center optimization methods for semiconductor manufacturing

Andreas Klemmt¹, Jan Lange, and Gerald Weigert

Electronics Packaging Laboratory
Faculty of Electrical Engineering and Information Technology
Dresden University of Technology
01062 Dresden, Germany

ABSTRACT
The planning and optimization of semiconductor manufacturing is a very complex task. Especially in the field of wafer processing - the so-called front-end - a lot of different processing steps have to be executed. These steps are for example typical batch tool operations like oven- and wet-etch processes, or typical cluster tool operations like dry-etch, implant or lithography processes which have complex setup strategies. Because of several specific constraints and dependencies, these steps are hard to schedule. Tools which perform equal processing steps are often organized as work centers. Normally, only dispatch rules are used to control and optimize these single work centers. In this paper, new optimization methods for the scheduling of processing steps are compared on the example of a representative oven work center. These methods are mixed integer programming (MIP) and simulation-based optimization. Mathematical models as well as simulation models for the selected work center are used to describe and to compare the different solution methodologies. All methods are investigated with respect to a later online usage. So partly hard time restrictions are added as optimization termination criteria. Furthermore, a complexity-focused analysis of the investigated problem is accomplished. This will show up to which limit of model-complexity mathematical methods are applicable. On the basis of a realistically generated manufacturing data set it is then investigated, which optimization potential can be exploited by the different methods for an oven work center. Thereby, input parameters, including the forecast horizon and their impacts on the objective weighted tardiness, are analysed.

1. INTRODUCTION AND LITERATURE SURVEY
The front-end wafer fabrication of a semiconductor manufacturing facility is special due to its numerous process steps, re-entrant flows, sequence-dependent setups and batch processing [1]. This is even more complex for facilities with additional concurrent operations like research and development processes. That means, a wider product mix and a potentially increased number of high-priority items. To reach a fast job cycle time the fab utilization is then typically lower than in a volume fab. Moreover, meeting customer due dates is one of the important manufacturing objectives. Because of the complex nature of the process, the customer due date is down-calculated on every operation as operation due date (ODD) [2]. The task is, to meet this ODD for every job in the work center as well as possible with respect to the job priorities.

This research focuses on the scheduling of furnace work center operations, which are commonly very long-running batch processes. Consequently, there is an extremely high impact on global manufacturing objectives [3]. We model the furnace work center operation as unrelated parallel batch processing machines with incompatible job families. Furthermore, we consider dynamic job arrivals and partly different job sizes. The optimization objective is total weighted tardiness. The problem is NP-hard.

Approaches to scheduling batch processes are a matter of particular interest and are also recorded in several publications. Hereby, the solution methodologies were grouped into exact approaches, approximate approaches and simulation approaches. A very detailed collection of these approaches is given in the literature review of [4]. Another review and classification can be found in [5]. In [3], dynamic programming and heuristic solution procedures are investigated for single machine batch problems with incompatible job families and objective total

¹ Corresponding author: Tel.: +49-351-463 31696; Fax: +49 351 463-37035 ; E-mail: klemmt@avt.et.tu-dresden.de
tardiness. This work was extended by [5] to the objective total weighted tardiness. In [1] and [6], the existing solution methodologies are extended to parallel batch machine problems. In [1], dynamic job arrivals are respected as well, which are essential for considering practical situations. Thereby, the problem is divided into two parts in almost all approaches: batching the jobs, and sequencing the batches on the machine(s). In [1] and [6], this decomposition is coupled with a Genetic Algorithm, in [7] an Ant Colony Algorithm is used. Unfortunately, for unrelated parallel batch machines, for example machines with different batch sizes and dedications, most of the decomposition approaches can not be used. In [8], MIP models as well as simulation models, containing these practical motivated process constraints, were developed. Here, the optimization objective was the job cycle time.

2. PROBLEM DESCRIPTION

In the following, an optimization approach for operative scheduling of a furnace work center is described. The work center consists of batch processing machines with partly different qualities. Related to [1] the problem can be described as follows (machine interruption is not allowed):

1. Jobs fall into different incompatible families that cannot be processed together.
2. There exist \( f \) job families. \( F := \{1, \ldots, f\} \) represents the set of all families.
3. There are \( n \) jobs to schedule. \( J := \{1, \ldots, n\} \) represents the set of all jobs.
4. There are \( m \) unrelated (non-identical) machines in parallel. \( M := \{1, \ldots, m\} \) represents the set of all machines.
5. The family of job \( i \) is represented as \( f_i \).
6. The (operation) due date of job \( i \) is represented as \( d_i \).
7. The priority weight for job \( i \) is denoted by \( w_i \).
8. The ready time of job \( i \) is represented as \( r_i \). 
9. The batch processing machine capacity of machine \( k \in M \) is represented as \( B_k \).
10. The dedication of machine \( k \) (allowed families on machine \( k \)) is represented as a subset \( D_k \subseteq F \).
11. The completion time of job \( i \) is denoted by \( C_i \).
12. The weighted tardiness of job \( i \) is denoted by \( w_iT_i = w_i \cdot (C_i - d_i)^+ \), where \( x^+ := \max(x,0) \)

3. SOLUTION METHODOLOGIES

In this paper three different solution strategies for the batch scheduling problems are investigated. Using the \( a | b | c \) notation of [9], the problems can be represented as: \( R_n | r_{batch, incompatible} | \sum w_iT_i \) and \( R_n | r_{batch, non-identical, incompatible} | \sum w_iT_i \) respectively. The solution strategies are mixed integer programming, simulation-based optimization and dispatching. All approaches will be described in detail in separate sections.

3.1. MIXED INTEGER PROGRAMMING

The development of the mixed integer program (MIP) follows [8]. Thereby some adjustments concerning constraints and objective become necessary. The parameter \( b \) restricts the number of batches on each machine. The subset \( M_b \subseteq M \) includes all machines which are allowed to process job \( i \). The subset \( J_b \subseteq J \) includes all jobs which are allowed for processing on machine \( k \) (both subsets could be directly derived from \( D_k \)). \( K \) is a significant large positive number. The following variables have to be defined:

\[
\begin{align*}
& s_{jk} \in \mathbb{R}_+ & \text{ ... starting time of batch } j \text{ on machine } k; (j = 1, \ldots, b; k \in M) .
\end{align*}
\]

\[
\begin{align*}
& x_{jk} \in \{0,1\} & \text{ ... job } i \text{ is scheduled in batch } j \text{ on machine } k, 0 \text{ otherwise; } (i \in J; j = 1, \ldots, b; k \in M) .
\end{align*}
\]

\[
\begin{align*}
& y_{jk} \in \{0,1\} & \text{ ... family } l \text{ is scheduled in batch } j \text{ on machine } k, 0 \text{ otherwise; } (i \in J; j = 1, \ldots, b; k \in M; l \in D_k \subseteq F) .
\end{align*}
\]

\[
\begin{align*}
& C_i \in \mathbb{R}_+ & \text{ ... completion time of job } i .
\end{align*}
\]

\[
\begin{align*}
& T_i \in \mathbb{R}_+ & \text{ ... tardiness of job } i .
\end{align*}
\]

As a next step, the basis batch process MIP-model has to be defined. Thereby, the constraints set (9) restricts the objective function (1). Constraints set (2) ensures that every job can only be processed in one batch on one machine. The maximum batch size is restricted by (3). Equations (4) and (5) force that only jobs of the same family can form a batch. Constraints set (6) guarantees that a job can only be processed after having reached the oven operation. Equation (7) forces that a batch can be started only after the last one is finished. Constraints set (8) restricts the completion time of every job.
Basic batch process MIP-model:

\[
\begin{align*}
\sum_{i \in J} w_i T_i & \rightarrow \min \\
\sum_{k \in M} \sum_{j=1}^b x_{ijk} & = 1 & i \in J \\
\sum_{i \in J} x_{ijk} & \leq B_k & j = 1, \ldots, b; k \in M \\
\sum_{i \in J} y_{jk} & = 1 & j = 1, \ldots, b; k \in M \\
y_{jk} - x_{ijk} & \geq 0 & i \in J; j = 1, \ldots, b; k \in M_i \\
x_{ijk} r_i & \leq s_{jk} & i \in J; j = 1, \ldots, b; k \in M_i \\
x_{jk} + p_{jk} & \leq s_{j+1,k} & i \in J; j = 1, \ldots, b-1; k \in M_i \\
K (1-x_{ijk}) + C_j & \geq s_{jk} + p_{jk} & i \in J; j = 1, \ldots, b; k \in M_i \\
C_j - T_i & \leq d_i & i \in J
\end{align*}
\]

Different modifications of this MIP-formulation are also possible. For example, it is not necessary to define the variable \(y_{jk}\) by using

\[x_{ijk} + x_{ijk} \leq 1 \quad i, h \in J; f_i \neq f_h; j = 1, \ldots, b; k \in M_i \cap M_h\] (10)

instead of (4) and (5). However, tests have shown that for the investigated batch problem this formulation is not as powerful as formulation (1) - (9). The number of binary unknowns is indeed lower than in (1) - (9) but the high number of additional constraints negate this advantage.

Additional process constraints:

In most practical applications several additional process constraints exist. The machine capacity for example often does not only depend on the number of jobs, but rather depends on the number of wafers the job consists of. Let \(W_k\) denote the maximum wafer capacity of machine \(k\) and let \(a_i\) the number of wafers of job \(i\). This means, we have to add the following constraints to the MIP approach:

\[
\sum_{i \in J} a_i x_{ijk} \leq W_k \quad k \in M; j = 1, \ldots, b
\] (11)

In the case of different job sizes the problem turns into \(R_{a,batch,non-identical,incompatible} | \sum w_i T_i\). Another practical relevant constraint is: not every machine is immediately available. Let \(r_k\) denote the ready time of machine \(k\), then we have to add the following constraint:

\[r_k \leq s_{i,k} \quad k \in M\] (12)

Also a job \(i\) can have a time coupling restriction \(t_i\), which defines the maximum allowed queuing time before starting the oven process (for example to avoid oxidation). Then the following process constraint has to be added for job \(i\):

\[s_{jk} - r_i - t_i + K x_{ijk} \leq K \quad j = 1, \ldots, b; k \in M_i\] (13)

In the following the two different variants of the MIP approach will be investigated: First (denoted as MIP-I) the problem is tried to solve directly with the derivated mathematical model. As we will see later, the solution quality decreases very fast with increasing problem complexity (section 3.4). Second (denoted as MIP-II) the problem is dissected and MIP-sub-problems were solved cyclic in the following way:

MIP-II: Every time \(t\) a machine becomes free, a MIP-sub-problem is solved. Thereby, only jobs with ready times lower than a time horizon \(t + \Delta t\) are considered. So, \(J := \{ i \in J | r_i \leq t + \Delta t \}\) is used instead of \(J\). Possible rest processing times on machines are set as initial down times. If no machine is available, the procedure is repeated at the next job release. Instead of (1) the objective function (14) is used for solving the cyclic MIP-sub-problems.

\[
K \cdot \sum_{i \in J} w_i T_i + \sum_{i \in J} C_j \rightarrow \min
\] (14)
3.2. SIMULATION-BASED OPTIMIZATION

In production control, if the complexity of practical problems is too high to be solved by exact methods, heuristic optimization algorithms in connection with simulation systems become suitable alternatives. Figure 1 shows the basic principle of a simulation-aided optimization system (see [10] for more details). The problem is described by a simulation model which includes a set of control variables and responds with an objective value $V$ after a simulation run is completed. The control vector $\mathbf{x}$ consists of several variables which influence the behaviour of the simulation model. So the optimization system is divided into an evaluation part on the one hand and a separate optimization algorithm on the other hand. Both subsystems communicate by $\mathbf{x}$ and $V$ while the optimization cycle is running. A heuristic search algorithm evaluates the objective function after a simulation run is completed and modifies the control variables. Then the model is simulated again (under slightly changed conditions). So the simulation-based optimization approach can be seen as a heuristically evaluated comparison of alternatives. Typically, iterative local search algorithms (Threshold Acceptance, Greedy, etc.) or Genetic Algorithms are used as optimization method. In this paper, an approach with a Genetic Algorithm (denoted as GA) is investigated. Three points of influence are chosen as control variables of the simulation-based optimization: First, in each case after a previous batch is finished, the selection of the next batch family and the corresponding jobs can be changed by the search algorithm. Second, a “minimum batch size” parameter is adjustable. This parameter defines that a batch starts immediately, if the specified batch size is reached. And third, the “maximum waiting time” can be varied. This value retards the start of a new batch, if additional jobs of the same family arrive in this period.

3.3. DISPATCHING

Four different dispatch approaches for scheduling batch processes will be investigated in this section. These are explicitly EDD, BATC, BATC-I and BATC-II which will be explained shortly in the following.

**EDD:** The EDD rule (Earliest Due Date [5]) is implemented in the following way: The jobs were ordered descending to their due dates. Every time a machine becomes free, a batch of the earliest job family (only an allowed family) is created within the capacity boundaries of the respective machine. If no machine is available, the procedure is repeated at the next job release. Note: This method completely ignores the weights and ready times of the jobs.

**BATC:** The modification of the ATC rule (Apparent Tardiness Cost [5]) for batch processes is implemented in the following way: Every time $t$ a machine $k$ becomes free, one batch $x_l$ from each allowed family $l$ is formed, and of all the formed batches the highest prioritized batch (concerning (16)) is chosen and scheduled on the machine. If no machine is available, the procedure is repeated at the next job release. It follows the ATC-priority calculation for every job $i$ of family $l$:

$$I_{il}(t) = \frac{w_i}{p_{lk}} \exp \left( -\frac{(d_i - p_{lk} - t)}{\hat{k} \cdot \hat{p}} \right)$$

The batch $x_l$, which contains highest prioritized jobs (concerning (15)) of each allowed family $l$, is created regarding to the machine $k$ and its capacity boundaries $B_k$ (or $W_k$). The batch priority of each allowed family $l$ is calculated by (16). Note: This method ignores ready times of the jobs.

$$BATC_l(t) = \sum_{i \in l} I_{il}(t)$$

**BATC-I:** The modification of the BATC rule (see [1]) for batch processes with ready times is related to the BATC procedure. In contrast to this, it follows the ATC-job priority calculation for every job $i$ of family $l$ at time $t$:

$$I_{il}(t) = \frac{w_i}{p_{lk}} \exp \left( -\frac{(d_i - p_{lk} - t + (r_i - t))}{\hat{k} \cdot \hat{p}} \right)$$

**GA-settings:** population size: 60, reproduction cycle: 15%, selection type: deterministic, crossover permutations: single point cross over, mutation: 3% (child’s only), mutation step size: 10%, random immigrants: after 4% equality
concerning (18) is chosen. Thereby the BATC-II (t) priority is calculated for all possible batch combinations of the first three jobs of each family (for more details see [1]).

\[
BATC-I(t) = \left( \frac{w_i}{p_i k} \right) \exp \left( -\frac{-d_i - p_i k - t + (r_i - t)^+}{k \cdot \bar{p}} \right) \min \left( \frac{n_{x_i}}{B_k^i}, 1 \right) 
\]

\[
BATC-II: \text{ The BATC-II rule is nearly identical to the BATC-I rule (see [1]). Instead of (18) the priority of batch } \ x_i \ \text{is calculated as:}
\]

\[
BATC-II(t) = \sum_{i \in \mathcal{J}} \left( \frac{w_i}{p_i} \right) \exp \left( -\frac{-d_i - p_i k - t + (r_i - t)^+}{k \cdot \bar{p}} \right) \min \left( \frac{n_{x_i}}{B_k^i}, 1 \right)
\]

4. BENCHMARKING AND RESULTS

In this section a first comparison of the different solution methods is introduced. For this, four representative example models from [8] are used (enlarged by \( d_i \) and \( w_i \)). All listed values are altered and do not allow conclusions of real manufacturing data. However, the problem dimension, like the number of investigated machines and jobs as well as the number of products and the number of allowed products per machine, are similar to usual production systems [8]. Thereby the initial problem OVE-1 could be solved exactly by the MIP-approach. So the problem dimension, like the number of investigated machines and jobs as example models from [8] are used (enlarged by \( B_k = 2 \) and \( 3 \) are the jobs of family \( f \) prioritized in decreasing order by (17). Then the batch priority (18) is calculated for all batch combinations \( \{4,3,1,7\} \). Optimality proof was only for \( b = 3 \) possible. Better solutions exists for \( b = 4 \) which are much harder to find.

Table 1: Example models OVE-1, OVE-2, OVE-3 and OVE-4 following [8]

<table>
<thead>
<tr>
<th>Problem</th>
<th>( n )</th>
<th>( m )</th>
<th>( f )</th>
<th>( B )</th>
<th>Problem colouring</th>
<th>Oven</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVE-1</td>
<td>25</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>OVE-1,-2,-3,-4</td>
<td>( W_i )</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>OVE-2</td>
<td>30</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>OVE-2,-3,-4</td>
<td>( r_j )</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>OVE-3</td>
<td>40</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>OVE-3,-4</td>
<td>( D_k )</td>
<td>1.2</td>
<td>1.2</td>
<td>4.5</td>
<td>1.6</td>
<td>1.2</td>
<td>1.2</td>
<td>5.8</td>
<td>9</td>
</tr>
<tr>
<td>OVE-4</td>
<td>60</td>
<td>8</td>
<td>10</td>
<td>4</td>
<td>OVE-4</td>
<td>( \bar{x}_{ij} )</td>
<td>3.4</td>
<td>3.4</td>
<td>6.7</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

| Job       | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| \( e_i \) | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2  | 3  | 3  | 3  | 4  | 5  | 5  | 5  | 5  | 5  | 6  | 6  |
| \( r_i \) | 1 | 1 | 2 | 1 | 5 | 5 | 5 | 5 | 3  | 1  | 3  | 1  | 2  | 2  | 1  | 1  | 6  | 6  | 6  | 6  | 6  |
| \( a_i \) | 25| 25| 20 | 25 | 25| 25 | 25| 25| 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| \( d_i \) | 11 | 10| 13 | 9  | 12 | 15| 14| 11| 17 | 12 | 15 | 16 | 18 | 16 | 18 | 16 | 22 | 21 | 21 | 20 | 24 |
| \( w_{ij} \) | 0.3| 0.3| 0.3 | 0.3 | 0.3 | 0.6 | 0.6 | 0.6 | 0.6 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.6 |

| Job       | 21| 22| 23| 24| 25| 26| 27| 28| 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|-----------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| \( e_i \) | 6 | 8 | 8 | 8 | 8 | 8 | 9 | 10| 10 | 10 | 10 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 | 12 |
| \( r_i \) | 6 | 1 | 4 | 7 | 7 | 1 | 1 | 3 | 3 | 3 | 8 | 8 | 8 | 6 | 1 | 1 | 7 | 7 | 7 | 7 | 7 |
| \( d_i \) | 25| 25| 26 | 23 | 30 | 22 | 25 | 30 | 22 | 26 | 22 | 26 | 28 | 28 | 27 | 29 | 27 | 28 | 28 | 28 |
| \( w_{ij} \) | 0.6| 0.3| 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 1  |

\( f \) are respected and the batch with the highest priority is chosen. Thereby the BATC-II (t) priority is calculated for all possible batch combinations of the first three jobs of each family (for more details see [1]).

\[ \text{BATC-I}(t) = \left( \frac{w_i}{p_i k} \right) \exp \left( -\frac{-d_i - p_i k - t + (r_i - t)^+}{k \cdot \bar{p}} \right) \min \left( \frac{n_{x_i}}{B_k^i}, 1 \right) \]

\[ \text{BATC-II: The BATC-II rule is nearly identical to the BATC-I rule (see [1]). Instead of (18) the priority of batch } \ x_i \ \text{is calculated as:} \]

\[ \text{BATC-II}(t) = \sum_{i \in \mathcal{J}} \left( \frac{w_i}{p_i} \right) \exp \left( -\frac{-d_i - p_i k - t + (r_i - t)^+}{k \cdot \bar{p}} \right) \min \left( \frac{n_{x_i}}{B_k^i}, 1 \right) \]
For all families identical processing times are assumed on all machines. The results of the different approaches are shown in Table 2. The termination criterion for the MIP-solver and the Genetic Algorithm was a time limitation of maximum 5 minutes, to simulate online conditions. For solving the MIPs, the CPLEX 11.0 library of TOMLAB was used. The simulation-based optimization approaches are processed by the discrete event simulation and optimization system simcron MODELLER. For GA, MIP-I and MIP-II, the average results of 10 repetitions are listed. Thereby, MIP-I consists of (1), (2), (4) - (9), (11), (12) and MIP-II of (14), (2), (4) - (9), (11), (12). The look ahead parameter \( k \), for the dispatch heuristics based on the ATC-rule (BATC, BATC-I and BATC-II), was increased from 0.5 to 5 in increments of 0.5 (similar to [3] and [6]). Then, the respectively best results are listed.

<table>
<thead>
<tr>
<th>Compare</th>
<th>EDD</th>
<th>BATC</th>
<th>BATC-I (thres = 10)</th>
<th>BATC-II (thres = 10)</th>
<th>GA</th>
<th>MIP-I</th>
<th>MIP-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVE-1</td>
<td>12</td>
<td>8.4</td>
<td>7.6</td>
<td>7.6</td>
<td>10</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>OVE-2</td>
<td>3.3</td>
<td>3</td>
<td>5.2</td>
<td>5.2</td>
<td>6.1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>OVE-3</td>
<td>9.5</td>
<td>9.2</td>
<td>8.5</td>
<td>7.7</td>
<td>9.4</td>
<td>13.1</td>
<td>14.6</td>
</tr>
<tr>
<td>OVE-4</td>
<td>37.2</td>
<td>11.2</td>
<td>16.6</td>
<td>19</td>
<td>16.6</td>
<td>17.8</td>
<td>17.8</td>
</tr>
</tbody>
</table>

As Table 2 shows, the EDD rule performs poorly, which has been also reported by [5]. Method BATC-I and BATC-II react sensitive on too high forecast horizons \( \Delta t \). The direct MIP approach MIP-I is unusable for problems bigger than OVE-2. Note: this approach strongly depends on the used objective function. In [8] the approach delivers good results up to problem OVE-4 for the objective \( \Sigma C_i \). The best results are reached by method MIP-II. Here, a high forecast delivers best results. But beware: The forecast influences the model complexity significantly (see OVE-4 with \( \Delta t = 9 \)). Continues good results were reached by the simulation-based approach.

Table 1 shows a benchmark which can be replicated by other researchers. First statements on the quality of the solutions methods could be made (Table 2). However, to substantiate the results, a statistical analysis for different assumptions has to be made. For this a benchmark is used, which is investigated in different variances by many researchers (see for example [1], [3], [5], [6], [7]). In the following (Table 3), this benchmark is adapted to the unrelated parallel machine batch problem. Thereby, some modifications have to be made to allow a modelling of practical conditions in more detail. So, to represent the high product mixture, the number of families is increased of 5 to 7. Furthermore, the work center utilization \( u \) is variable from 0.6 to 0.9. Here, an average batch size \( b = 0.75 \) is assumed. Let \( \hat{p} \) denote the average job process time, then the expected makespan of each benchmark follows:

\[
\text{makespan} = n \cdot \hat{p} \cdot \left( \frac{\hat{b}}{u} \cdot \sum_{k=1}^{m} \beta_k \right)^{-1}
\]

Table 3: Benchmark generation - Test date parameters

<table>
<thead>
<tr>
<th>Problem parameter</th>
<th>Values used</th>
<th>Total values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of machines</td>
<td>3, 4, 5</td>
<td>3</td>
</tr>
<tr>
<td>Number of jobs/family</td>
<td>20, 30, 40</td>
<td>3</td>
</tr>
<tr>
<td>Number of families</td>
<td>5, 6, 7</td>
<td>3</td>
</tr>
<tr>
<td>Utilization</td>
<td>0.5, 0.6, 0.7, 0.8</td>
<td>4</td>
</tr>
<tr>
<td>Family processing time( t )</td>
<td>2 (20%), 4 (20%), 10 (30%), 16 (20%), 20 (10%)</td>
<td>1</td>
</tr>
<tr>
<td>Weight per job</td>
<td>0.3 (75%), 0.6 (20%), 1 (5%)</td>
<td>1</td>
</tr>
<tr>
<td>Release date</td>
<td>( r_i = \text{Uniform} (0, \text{makespan}) )</td>
<td>1</td>
</tr>
<tr>
<td>Due date</td>
<td>( d_i = \text{Uniform} (r_i - (v - 1) \cdot \hat{p} + r_i + (v + 1) \cdot \hat{p}) )</td>
<td>1</td>
</tr>
<tr>
<td>Batch size</td>
<td>( B_1 = 4, B_2 = 4, B_3 = 5, B_4 = 2, B_5 = 2 )</td>
<td>1</td>
</tr>
<tr>
<td>Dedication (allowed families)</td>
<td>( D_1 = {1, 2, 3, 4}, D_2 = {2, 3, 4, 5}, D_3 = {5, 6, 7}, D_4 = {1, 6, 7}, D_5 = {2, 3, 5} )</td>
<td>1</td>
</tr>
<tr>
<td>Problem instances</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

\( ^1 \text{CPLEX-Settings: NODEALG = 1; SUBALG = 1; PREIND = 0; AGGIND = 0; all cuts disabled} \)
\( ^2 \text{The processing time of a family is assumed to be equal on all machines} \)
The ready times of each job are uniformly distributed within the interval \([0, \text{makespan}]\). Assumed that all jobs are controlled by ODD in pre-operations, the due-date spreads around the ready date. Here, a deviation of five \((\nu = 5)\) average process times is assumed. The weight of the jobs is not uniformly distributed, to represent the different priority items. In contrast to other researches, the batch size is not variable but fixed at different capacities for the machines. Also, an additional dedication is attached. Altogether 1080 different benchmarks, with job ranges from 100 to 280, are investigated. To keep the computational effort in acceptable boundaries, only the method MIP-II as well as the dispatch approaches EDD, BATC, BATC-I and BATC-II, are investigated further. Because of the stochastic evaluation of the family processing times, not only the fixed forecast horizons \(\Delta t = 4\) is investigated for the methods BATC-I, BATC-II and MIP-II, but also \(\Delta t = \hat{p}\). The threshold parameter in method BATC-I and BATC-II is set to \(\text{thres} = 10\). For the MIP sub problems a maximum optimization time of 30s is set.

The results of this practical motivated benchmark are shown in Table 4. Altogether 8640 different schedules are created with the different solution methods discussed. The table is to read as follows: In every row it is displayed the average total weighted tardiness value of all benchmarks with their respective method and their denoted compare item (e.g. \(m = 3\)), normalized on the EDD solution. So different methods become comparable by a static coverage and the objective function improvements can be directly read off (compared to EDD). As Table 4 shows, most of the statements reached in the OVE example models can be verified. So, the cyclic MIP approach also reaches the best results in this benchmark. Assuming that \(\hat{p}\) is greater than 4 hours in general, it can be verified that a higher forecast horizon leads to better results for the MIP approach. The influences of the parameter settings on the quality of the MIP solution can be summarized as follows: An increased number of machines lead to an increased problem complexity, so the solution quality decreases. The number of jobs has no influence (because of the cyclic planning). The influence of \(f\) is balanced: A higher number of families increases the number of batches in each sub problem in average but it decreases the number of possible job combinations in the batches. With an increasing utilization the optimization potential increases. If the problem complexity is not too high, this leads to better solver results. The results of the dispatch approaches can be summarized as follows: On average BATC, BATC-I and BATC-II reach better results than EDD. But as mentioned before, here the forecast horizon should be chosen tightly. Overall, method BATC-II performs better than BATC-I, especially in the case of high utilization. The reason for this lies in the characteristic of the BATC-II rule to increase the fullness of the batches.

In the case of a benchmark, the ready dates of all jobs are known. So the problem is tried to attack as a hole by many researches. For practical application, it has to be precisely proven, what forecast horizon \(\Delta t\) really exists. So often (i.e. for short forecast horizons), only time window approaches like BATC-I, BATC-II or MIP-II can be then used in an online application. As Table 4 shows, the different methods react very sensitive on \(\Delta t\). If the forecast horizon turns to 0, there is only a small optimization potential for the MIP approach. In contrast, with a forecast around \(\hat{p}\), the optimization potential is strongly increased. But beware: In this case the model complexity also increases. Here, it has to be investigated the exact character of the boundaries (e.g. the number of jobs in the MIP-sub-problem) specific to the work center. In the benchmark, the average (maximum) number of jobs in each MIP-sub-problem for the smallest problems \((m = 3, \ u = 0.6)\) was: 12 (27), and for the largest problems \((m = 5, \ u = 0.9)\): 20 (42). So, small problems could be solved in a few seconds. However, the optimization of the large problems takes 30s for nearly each batch calculation in every optimization cycle.

<table>
<thead>
<tr>
<th>Compare</th>
<th>Machines</th>
<th>Jobs per family</th>
<th>Families</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(m = 3)</td>
<td>(m = 4)</td>
<td>(m = 5)</td>
<td>(j = 20)</td>
</tr>
<tr>
<td>BATC</td>
<td>0.894</td>
<td>0.926</td>
<td>0.959</td>
<td>0.929</td>
</tr>
<tr>
<td>BATC-I</td>
<td>0.940</td>
<td>0.948</td>
<td>0.971</td>
<td>0.941</td>
</tr>
<tr>
<td>(\Delta t = 4)</td>
<td>1.220</td>
<td>1.193</td>
<td>1.168</td>
<td>1.117</td>
</tr>
<tr>
<td>BATC-II</td>
<td>0.894</td>
<td>0.914</td>
<td>0.945</td>
<td>0.916</td>
</tr>
<tr>
<td>(\Delta t = \hat{p})</td>
<td>1.076</td>
<td>1.068</td>
<td>1.073</td>
<td>1.038</td>
</tr>
<tr>
<td>MIP-II</td>
<td>0.853</td>
<td>0.888</td>
<td>0.920</td>
<td>0.891</td>
</tr>
<tr>
<td>(\Delta t = 4)</td>
<td>0.839</td>
<td>0.875</td>
<td>0.905</td>
<td>0.875</td>
</tr>
<tr>
<td>(\Delta t = \hat{p})</td>
<td>0.839</td>
<td>0.875</td>
<td>0.905</td>
<td>0.875</td>
</tr>
</tbody>
</table>
5. FURTHER INVESTIGATIONS

Different modifications of the MIP model will be investigated in future. For example the solution quality strongly depends on parameter $b$. Here different heuristics are developed to set this parameter sensible. It is possible to restrict $b$ machine specific in dependence to the actual job list and its dedications. Furthermore a job list reduction of $J$ has to be investigated in the case of too high job stocks, to reduce model complexity. For highly distributed processing times it could be sensible to adapt a set of available machines and use $M := \{ k \in M \mid r_j < t + \hat{p} \}$ instead of $M$. In theory it is also possible, to get a strong upper bound for the objective function of the MIP. Here for example the BATC rule could be used to create an initial solution for the solver. The goal is an adaptation of the investigated solution methods on real manufacturing data sets (as done in [8]).

6. SUMMARY

The present investigations are based on commonly known benchmark models. The oven group of a semiconductor front-end process is only an (but well investigated) example for several batch processes in industrial practice. It has shown that for practical use the (weighted) tardiness is much more important than the makespan or other objectives. Some additional constraints, such as equipment dedication or different batch sizes, have to be considered. On these conditions a MIP approach for scheduling was suitable only for small problem sizes. In contrast, heuristic simulation-based algorithms are not this sensitive to the objective type and keep applicable also for larger problem dimension. But often they fail in practice due to their high time consumption.

But our experiences in operational scheduling teach us that the problem dimension can be mostly significantly decreased by shortening the time horizon. Inside these shorter time slides a MIP optimization is possible as well as simulation-based heuristic algorithms. This approach of a quasi time continuous optimization comes closer to the practical requirements of an actual process accompanying scheduling than most of the theoretical benchmark models.

REFERENCES

STRATEGIC RESOURCES MANAGEMENT INTO HEALTH CARE SYSTEM: THE ED CASE

Fabio FRUGGIERO1*, Alfredo LAMBIASE2 and Luca ZAMPOGNARO2

1Dipartimento di Ingegneria e Fisica dell’Ambiente
Università della Basilicata
C.da Macchia Romana, via Ateneo Lucano, 85100
Potenza (Pz) - ITALY

2Department of Mechanical Engineering
University of Salerno
via Ponte don Melillo 1, 84084
Fisciano (SA) - ITALY

ABSTRACT

An operational capacity management, in most acute hospitals, could aim to evaluate system features according to utilization rate, delivery time and costs analysis. Moreover, there is a significant potential into application of computer simulation tools in predicting and managing overcrowding and long waiting time. In the proposed network an optimizing multi-objective approach has been adopted for patients and staff scheduling problem. Performing cost’s analysis into treatments and policies, evaluating patient’s throughput and waiting time authors look for a complete system portrayal and evaluation of alternative management strategies. Linear programming, priority rules, timed queueing networks, jockeying from one queue to another, facilities layout have been treated to carry out the simulated environment. System flexibility, service level to patient, real time reaction to unexpected events have been analysed. The aim of the present work is to investigate, and improve, healthcare performances and scalability in a real system analysis (i.e., the Emergency Department environment). Simulation approach and system and patient performances analysis have been merged in order to support planning alternatives. Timed Queueing Petri nets, Fuzzy logic and constraint programming have been used into modelling treatment time and flow path definition. An Analytic Hierarchy Process (i.e., AHP) has been used into the decision making process.


1. INTRODUCTION

Healthcare system represents a complex business. It has to deliver patients to a timely and effective medical assistance. National government spends any amount of money with the attempt of supplying efficient service level to patients. Cost is inarguably one of the central features in every discussion of the healthcare delivery system and/or every potential replacement and/or reform (i.e., medicare’s revenues and expenses continue to rise every year entailing a financial risk of longevity [1]-[2]).

Health care assistance is mainly provided through medical centres and hospitals. They offer care to people who are often at their vulnerable point in life joining expertise knowledge, lifesaving technologies and drug cure. Cost containment isn’t about cutbacks but rather better strategy, better management, better service provision, competitiveness and taking the correct options [3]. In the meanwhile the 10% of hospitals in the United States had closed for problems associated with poor management skills [4]-[5]. Frequently the amount of money spent into healthcare has been used just like a knee-jerk reactions to a medical mishap or to perceived crisis.

Operation management techniques and best practices should be used as a mainline for future crazes and improvements [6]. It has been recognized there is a significant potential into application of computer tools in predicting and managing overcrowding and long waiting time [7]-[8]-[9]. Investigation into real emergency facilities is going to be carried on. Computer simulation, visualization, can be used to test system’s behavior anticipating the potential performance of the system in different hypothetical arrangements [10]-[11]. Improvements/balancing into service level have been manifested [12].
The behaviour of the implemented approach is the Emergency Department (i.e., ED) of the Cork University Hospital (CUH) – Cork – IRELAND. In an ED, because of its complex feature (i.e., queueing problems, preemptive priorities) traditional operating research techniques are of limited usefulness [13].

The authors made use of Timed Queueing Petri Nets approach [14] into modelling system’s behaviour and analyzing performance and scalability. Moreover another important aspect is the possibility to gain a balanced schedule, to simulate the starting time, and the rota as well, according to patient’s arrival rate. Because of the manifold number of possible configurations, it has been decided to apply a heuristic procedure, based on neural networks, into optimization perspective. The optimization in beds and resources capacity planning was implemented in order to explore the impact of unplanned admissions on patients turnover, staff adequacy and system productivity. A decision making procedure was used into costs weighing up mechanism.

2. DISCRETE EVENT SIMULATION INTO EMERGENCY DEPARTMENT (ED) OF CUH

The Emergency Department (i.e., ED) of Cork University Hospital currently services around 50,000 patients a year (data were “Poissonly” distributed as after admission office registration – see fig. 1). The geographic collocation of the hospital is going to serve a catchment population of approximately 300,000 people who wish to receive a timely and effective medical assistance. The ED provides quite direct medical tests for patient and supports hospital’s mission. Mainly the ED is the 24 hours inner patients gate to medical facilities (i.e., the 70% of in-hospital population comes through ED into CUH case). This places a heavy burden on the facilities, space, staff and finances of the hospital.

The layout of inquired ED includes reception area, a separate triage area to assess patients, a waiting hall and treatments areas (there are separate adult and child rooms). Four places resuscitation cubicles with overhead x-ray equipment are furnished into ED facilities. These are intensive care units for medicine and surgery of critical ill patients. Moreover, there are two x-ray examination rooms, 12 major bed clinical decision units, an emergency eye treatment facility. Treatment areas for minor injuries have been set up. Besides, the system is provided with special cardiac care equipments, renal and blood test equipments (like major treatment services), plaster and suture facilities and skilled nursing for long term care (as minor room).

The layout of ED in CUH was investigated using the Quadratic Assignment Problem (i.e., QAP) [15] formulation to validate the location/distance of the set of treatments room based on flow paths and constraints [16]. Through a triage process patients are categorized, encoding in Manchester emergency rules [17], in 5 categories (i.e., queues) or patient’s lists. The queues are organized as set of analogous pathologies (i.e., jobs). Each pathology is cured performing different chains of treatments (i.e., tasks) policies. Precedence constraints (i.e., fixed sequences of treatments), zoning constraints (i.e., forced assignment of treatments to resources, e.g. specialist consultation), capacity constraints (i.e., limited max saturation rate) tighten the tabu list of possible choices. The daily arrival trend can be fixed across a Poisson function (eq. 1) distribution.

\[
Pr(X(t)=k) = \frac{(\delta t)^k e^{-\delta t}}{k!}
\]

**Figure 1**: The “Poissonly” annual patients distribution by time strip

Where \(\delta\) is the shape parameter (different from pathologies) which indicates the average number of events in the given slot \([0,t]\). Daily and nightly trends, and growth of arrival during the bank holidays (an increment of 15%), were noticed. The time strip between 11 a.m. to 20 p.m. embodies on average the 65% of patient arrivals.

The ED simulation model has been built as a set of interactive components described across timed Queueing Petri Nets [14]. Such a network is able to prospect all possible paths into the system. A jockeying policy, based on treatment waiting time and weighted on pathologies, has been implemented across queues. Admission of patients
can be facilitated if certain tasks are, and/or need, handled prior to the patient being admitted to the hospital. The main streamlines into the ED simulation model follows immediately emergency admissions patients (i.e., care type 1 and 2), urgent (i.e., care type 3), elective (i.e., care type 4 and 5). The simulation model is organized as a pull claim.

2.1. THE ED MODELLING PHASE

A traditional view of the emergency department is as that of drain on hospital resources. It is mainly considered as the access block to the hospital facilities. The flow of patients through a hospital ED is a complex dynamic and fluid process, affected by multiple factors external to ED (such as patients arrival rate and pathologies diagnosis). Different steps have been necessary to learn, and consequently characterize, the ED environment, procedures and paths into a simulate environment. First of all, patient counselling characterization (into duties and time), treatments zone and staffing activation follow up a place/transition network. Processing sequences are mainly based on priority rules (basically organized as FIFO). Expert opinion (i.e., managerial and hospital staff), as counsels and questionnaire (mainly in time and cost characterization), has been worked out. Tailing on discrete event simulation, state charts can work out the scheduling procedures. Moreover, a Petri Net is going to describe synthetically the manifold patient’s paths.

The Petri Net (i.e., PN) is a place/transition network which defines a graphical, executable technique for the specification and analysis of concurrent, discrete event system [18]. This framework provides a set of building blocks for specifying the state transition mechanism. Each of the several events associated with a discrete event state competes to trigger the next state transition. At each state new events may be scheduled. The network as in Petri (i.e., a 4-tuple \( P=PN(P,T,F,p_0) \)) is specified by a finite set of places (i.e., \( P \)) and a finite number of transitions (i.e., \( T:\mathbb{P}\cap T\neq\emptyset \)) along with inhibitor input/output functions (i.e., \( F^+, F^- \) as backward and forward incidence functions \( F \subseteq (P\times T)\cup(T\times P) \rightarrow \mathbb{N}^+ \), the flow relation paths) and initial state configuration (i.e., \( p_0 \in P \)).

![Figure 2: The ED routing module as in Timed Queueing Petri Nets modelling](image)

A marking is an assignment of token (i.e., a count or rather a patient) to place of the net. In addition, the enabling/disabling mechanism of the arcs can be fired (i.e., based on emergency paths). Timed transitions (according to observation and consultant/doctor suggestions) mainly based on fuzzy membership functions [19], simulate the processing time. As a graphical tool (i.e., a bipartite graph in which circles represent places, bars represent transitions and arrows represent the flow relation), the PNs can be used like a visual-communication aid similar to flow charts and networks [20]. Inasmuch as the simple PNs are mainly a qualitative description of the systems behaviour, combining them with Queueing Networks (i.e., QNs) let to further assess the quantitative performance properties (fig. 2).
A token arriving at a queue will immediately be served if a free server can be allocated to him. Pre-emption is disabled. A token added to a timed place is inserted into a queue (as after the Manchester triage process) according to the queue’s scheduling strategy (eq. 4). The process starts with a queued registration and a timed queueing dispatch procedure (i.e., triage state). Patient on this “place” are available for transitions (i.e., cure state) if there are unoccupied resources/staff. The class j corresponds to a pathology as after the triage process (the 80% of possible paths correspond to i=20). After main treatments the patients can be hospitalized or discharged.

2.2. A SIMULATED DECISION MAKING ENVIRONMENT

According to the plethora of possible paths and situations, the usefulness of the implemented discrete event simulation model is twofold. First of all the model “simulates” a Decision Making (i.e., DM) support to managerial strategies [21]. The simulation foretells the behaviour of ED amenable to dimensional changes into staff (i.e., No. of nurses as well doctor and technicians etc…) and environment (i.e., No. of major/minor cubicles, resuscitations rooms and so on) and performances. Of far more relevance is the use of simulation models to initiate changes within the emergency facilities and procedures while improving conditions with much less costs. Secondly, the implemented simulation model renders, if opportunely framed into the real environment (such a frame was implemented using java commands and Visual Basic modules), a “hot dynamic interface” to timely system’s optimization.

An Analytic Hierarchy Process (i.e., AHP) methodology (i.e., a hierarchically Multi-Criteria Decision Making, i.e. MCDM, procedure) helps determining the relative importance of each alternative in cure [22]-[23] (i.e., on selecting relevant resources and facilities, on defining treatments, engaging hospital staff, consumption of medicines etc…). Medical (i.e., managerial and doctors) counsel were engaged as expert opinion. The pairwise judgments were arranged in matrices. An index consistency (based on eigenvalues analysis) measures the inconsistencies in judgments.

Managing the inner patient level, seeing the changing demand and facility and resources, the simulation model is going to measure resources saturation and time to service reply. A planning perspective is going to drive into a more competitive environment seeing the proper long-range overall ability to deal with future and contingencies. The authors recognized as an important parameter the investigation of the reaction time of the system. It represents the competitive environment seeing the proper long-range overall ability to deal with future and contingencies. The simulation can point out bottlenecks. These are the access blocks of the ED system (e.g., the patient remains in the ED, as a cost and visual inefficiency value, until an inpatient bed/resource can be allocated).

2.3. ED OPTIMIZATION IN BALANCING PERSPECTIVES

The optimization of resources into ED consists in finding an optimal configuration in order to perform the scheduling of patients (i.e., jobs: \( J = \{ J \}_{i=1}^{m} \)) to resources (i.e., places: \( W = \{ W \}_{i=1}^{n} \)).

Each patient is characterized by a Release-time (i.e., \( R_{i} \), the arrival time into the ED according to eq. 1), a priority-level (i.e., the triage care type \( L \)) and a medical-route (i.e., treatments to attend the patient’s disease \( T = \{ T \}_{i=1}^{n} \)).

Each patient \( J_{i} \) can be “processed” on interchangeable resources (compatible with processing, zoning and capacity constraints) and requires a medical-route, i.e., a chain of tasks \( O_{i1}, O_{i2}, \ldots O_{im} \).

Each task of the medical-route is performed in a fuzzy \( t_{ik} \) time. \( O_{ik} \) is the treatment of patient \( J_{i} \) that has to be process on medical resource \( W_{k} \) for a treatment time \( t_{ik} \). As various alternative runs were made, the number of staff, and the rota starting time, was increased or decreased to gain an overall utilization in the 70% to 80% range for each shift. If \( h \) is candidate to be the assigned to resource \( k \), the workload on this resource is formulated across the following statement (eq. 2). Zoning and capacity constraints need to be satisfied.

\[
W_{i}(t) = \sum_{i \in X} \sum_{j \in L} t_{ij} + t_{ik} \quad \forall W_{i} \in W
\]

Where \( R_{ik} \) is the set of tasks (i.e., examinations) already assigned to resource \( k \) and \( t_{ij} \) is the processing time as after the defuzzification process and \( i \in \mathbb{N} \) is the No. patients into the system at time \( t \). A simple balancing procedure (eq. 4) is sought in machine selecting [24].
Based on constraints and resources availability each patient is characterized by a time of arrival $II_i$ (formally expressed into mm/dd/yy h:mm) and an exit time $TT_i$, if $R$ is the total number of tasks/treatments to cure disease $x$ related to patient $i$, LOS$_i$ is the difference between $II_i$ (e.g., based on the signing of emergency declaration, e.g., hospitalized or discharge or other hospital cure) and $TT_i$ in the form of (mm/dd/yy h:mm), as counter of the simulation. A processed time ($RTime_i$) has been evaluated (eq. 3).

$$RTime_i(t) = \sum_{j=1}^{\infty} TT_{a_j} - S_{a_j} \quad \forall j \in J, \forall T_i \in T$$

Where $T_i$ is the set of tasks assigned to job $J_i$ and $S_{a_j} \geq 0$ (i.e., the starting time of $k$-th task/operation of $i$-th job/patient). Based on these considerations it is possible to define the objective function $C^*$ every work shift.

$$C^* = \min \left\{ \alpha \left( \frac{1}{N} \sum_{i=1}^{N} LOS_i - RTime_i \right) + \beta \left[ \frac{1}{M} \sum_{k=1}^{M} \left( \sum_{i=1}^{N} OP_{a_k} - W_{a_k} \right) \right] \right\}$$

(4)

Where:
- $\alpha, \beta$ represent the weights of the objective function. High importance of waiting time responds to patients’ perspectives. The rising up of saturation rate responds to managerial perspective;
- $OP_{a_k}$ represents the operating time of resource $k$ of model $i$;
- $N$ is number of patients into the system at time $t$;
- $LL$ is the number of resources of the same type into the system;
- $M$ as the number of resources (i.e., medical staff) into the ED at time $t$.

The performance of the system has been, timely, reported according with saturation rates and the idle time of resource type $k$ (i.e., $SS_k$) and overall system (i.e., IT).

$$SS_k = \frac{1}{LL} \left[ \sum_{i=1}^{M} OP_{a_k} - W_{a_k} \right]$$

(5)

$$IT = \frac{1}{M} \sum_{i=1}^{M} SS_i$$

(6)

The $h$ candidate is assigned to resource $k$ in relation to resource’s availability at time $t$ and saturation rate of interchangeable resources (i.e., some treatments can be performed by nurse and/or porter, doctor and/or specialist etc…). When the task can be performed by parallel workstations the objective of the optimization procedure is to balance the system picking from idle resources.

3. PERFORMANCES ANALYSIS IN ED OF CUH

A general heuristic approach was used in order to guarantee exploration and exploitation of domain in optimal solutions (eq. 4). A tool (i.e., OptQuest Engine) based on meta-heuristic, mathematical optimization and neural networks components [25] was used. The optimized configuration was compared with the current ED arrangement of staff (i.e., 13 nurses, 2 porters, 6 SHOs – Senior House Officer, 4 registrars, 2 specialists). Moreover, an equipment analysis were performed. The outputs were expressed in terms of Reaction Time (i.e., the mean waiting time for the first treatment), total Waiting Time and the Length of Stay of patients into the ED facilities. A reduction of about 25% was prospected in optimal case.

Figure 3: Alternative ED system configurations. (a) Resources saturation rate. (b) Reaction time, Waiting Time and Length Of Stay in current and optimal ED configuration.
The balanced formulation was extended into rota definition. According to daily and weekly patients trend of arrival it is possible to coordinate the configuration and workshift (a slack of 2 hours per shift was simulated). Day, late and night workshifts were defined into slot and size. “Earlies” and “latest” was defined primarily as improving of customer service level. In the “7 day shift” optimal pattern, the employees can work their allotted hours with 0.65 mean saturation rate without spoiling the LOS for all patients. A weekly average reduction of 2 nurses and 1 doctor can be obtained with optimal scheduling procedures. Moreover, the analysis of ED performances is obviously related with the cumulative hospital wards capacity.

The bed assignment was simulated. Pointing out the increase of beds (i.e., 120→150→200 units) the aisle stay time in ED area was reduced (i.e., 280→15→10 minutes) with cost containments (i.e., 83.50→10→8.50 Euros) as staff and resources employment redundancy. A nurse isle was needed as the bed capacity was reduced under 150 units. An analysis of cost per time unit (0.8 euro/min) and main pathologies (detailing exams and drugs and staff cure – fig. 4) was simulated into the ED process by means of the AHP procedure. More generalized is the symptomatology and higher are the treatments/tests and costs. This analysis can manifest a ticket investigation.

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>€</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limb Problems</td>
<td>40</td>
</tr>
<tr>
<td>Abdominal Pain</td>
<td>80</td>
</tr>
<tr>
<td>Back Pain</td>
<td>25</td>
</tr>
<tr>
<td>Chest Pain</td>
<td>660</td>
</tr>
<tr>
<td>Neck Pain</td>
<td>36</td>
</tr>
<tr>
<td>Head Injury</td>
<td>80</td>
</tr>
<tr>
<td>Headache</td>
<td>45</td>
</tr>
<tr>
<td>Unwell Adult</td>
<td>23</td>
</tr>
<tr>
<td>Short Breath</td>
<td>44</td>
</tr>
<tr>
<td>Wounds</td>
<td>32</td>
</tr>
</tbody>
</table>

Figure 4: Simulated costs analysis for main treatments according to an AHP decision making process.

4. CONCLUSIONS

This study has examined the improvement of ED functions through the aspects of capacity efficiency and coordination.

A discrete event simulation model of healthcare behaviour creates the right condition for a continuous improvement system in order to test and evaluate alternative configurations, layouts, hypothetical settings, costs and procedure. Inasmuch as the system is complex in nature, the modeling phase was, with no doubts, the most complex process.

The impact of resources availability/efficiency/configuration, rota shifts, bed assignment were examined in overall systems performances. A simple model characterization with TQPNs were prospected. Visualization of procedures and scheduling alternatives were drafted. The main patients’ waiting time, Length Of Stay and system response to in patient’s arrival were checked in different and optimal arrangements. A real time link with unexpected events was performed according to a useful mechanism/procedure of input selecting.

ACKNOWLEDGEMENTS

The authors are thankful to the staff of Cork University Hospital and Cork Institute of Technology for their professional support, providing facilities and material for this research.

REFERENCES


Solving Complex Crew Allocation Problems in Labour-Intensive Industries Using Genetic Algorithms

Nashwan Dawood¹, and Ammar Al-Bazi

Centre for Construction Innovation Research
University of Teesside
Middlesbrough, TS1 3BA, UK

ABSTRACT

The high cost of skilled labour in the precast concrete industry and the dynamic nature of the production processes encouraged senior managers in the industry to develop more intelligent and optimal labour allocation strategies. In this paper, a Genetic Algorithm (GA)-based simulation optimisation approach is used for optimal allocation of different crews of workers based on different precast concrete production processes. A process simulation model is integrated to a GA-based optimisation model in order to simulate the physical processes that are involved in a labour-driven facility and to optimise the allocation of labour crews in that facility. The outcome of the proposed approach determines the optimal or near optimal allocation of crews to labour-intensive processes. This should eventually lead to maximum utilisation of a set of skilled workers involved in the allocated crew and subsequently minimise total labour costs. The paper discusses a simulation system dubbed “SIM_Crew” developed during the study. The simulation model is developed initially as a test bench for the proposed allocation system. A GA is used to guide the simulation towards the best course of action; with the chromosome designed to consider all of the decision variables. A probabilistic selection procedure has been developed in order to guarantee various selections of chromosomes. A sleeper precast concrete is developed as a case study to prove the proposed allocation concept. The results showed that efficient utilisation of skilled labour has a substantial impact on reducing throughput time, minimising labour costs, idle times and maximising the skilled workers utilisation.

1. INTRODUCTION

Many of the labour-intensive facilities involve highly skilled workers and experienced supervisors are required to carry out the processes. Labour intensive production processes are those processes that require a high level of labour compared to capital investment. Such processes are more likely to be seen in job shop production scenarios. The high cost of skilled labour in labour-driven processes is being a force driving the production planner to improve productivity and hence minimise the total production cost. To improve the productivity of a production facility, optimal allocation of resources and optimal planning of workforce is required. The best allocation of resources will eventually lead to minimising waste and guarantee the best flow of work. Production managers strive to achieve best allocation of resources on different processes. This is a complex problem due to the large range of different possible solutions. Traditional solutions to this type of problem do not produce satisfactory results. More sophisticated tools are required to assist in the decision making processes involved in resource allocation. The lack of utilisation of innovative tools to solve crew allocation problems in labour-intensive industries motivated this research aimed at assisting production planners as a decision support system in order to improve performance of their facilities. Optimal allocation of crews will eventually minimise labour costs, maximise labourer utilisation, minimise process idle time and subsequently improve the productivity of labour-intensive processes. The paper is structured as follows: in section, problem definition is demonstrated. In section 3, literature review outlining crew allocation techniques is presented. In section 4, SIM_Crew modules are addressed in detail. The materials and methods, with the used in designing the allocation model, is discussed in details in section 5. A case study demonstrated in section 6 followed by experimental activity in section 7. Section 8 addresses the analysis and interpretation of outputs. The conclusions and future work is presented in section 9.

¹ Corresponding author: Tel.: (+44) 1642- 342 494, E-mail: n.n.dawood@tees.ac.uk
2. Crew Allocation Problem in Labour-Intensive Industries

The problem is associated logical dependencies and resources (e.g., crew of workers and equipment such as machines) required when allocating resources to a labour-intensive project in a job shop environment. The formation of a crew for example, may contain some shared workers, involved in more than one process. The required crew for each process may be allocated individually or may come from other process activities that share the same resource/resources. The individual resource allocation allows activities to start as soon as the logical dependencies (i.e., the completion of the preceding activities) are available, with no delay (idle time for the resources). Since both the parallel and sequenced activities of some or all production processes of a project is pre-specified, the labour shared by multiple similar or different processes are generally allocated according to the work flow path. However, idle times or resources are may be not fully considered and the overall throughput time maybe longer or total costs a higher than expected. This type of problem becomes more important when there is significant idle time for a process or processes. This is caused by shared workers being allocated to more than one process and being required at different times dictated by the sequence requirements of similar labour-intensive operations. In order to increase productivity and reduce costs, optimal crew allocation in the labour-intensive facility is needed. The correct sequence application of each allocated worker to one or more than process has to be achieved in order to minimise the idle time caused by shared usage of an individual worker. The productivity and efficiency of manufacturing systems of this nature will be improved by the simulation system being developed.

![Figure 1: The problem of crew allocation on similar/different processes](image)

In figure 1, each parallel processes have a certain pool of crews that can possibly perform the work for each of the parallel process, each crew has a different set of workers with a certain process time. The formation of each crew (set of workers involved) may contain shared labour. When the shared labour available is split into more than formation it will eventually cause a delay in process/processes.

3. Literature Review for Crew Allocation System

Several research projects using innovative tools to improve the performance of the precast concrete product manufacturing systems have been conducted. Reference [1] combined Tabu search principles with a simple improvement- swapping heuristic to allocate stands and cutting patterns to logging crews for a single time period. Two comparisons were made between the TABU heuristic (500 iterations) and an Integer Programming (IP) formulation for a small problem (10 stands, 5 logging crews, 5 cutting patterns, and 5 log types). The first test allowed as many logging crews to be located in any one stand as was required. The Tabu search solution was within 0.8% of the IP solution. The second test allowed a maximum of one crew per stand. Reference [2] introduced a discrete event simulation for crew assignment and crew movements because of train traffic, labour rules, government regulations and optional crew schedules. The system helps to evaluate changes to current crew assignments and can test new crew assignment scenarios such as crew schedules. The system is also used to assess the impact of traffic changes on existing crew schedules in order to implement reactive corrections to these schedules. The results of the simulation allow the user to draw conclusions concerning the operational characteristic of the real crew assignment process. Reference [3] proposed a new methodology that utilises and combines Genetic Algorithms and spatial technologies for optimisation of crew formations for earthmoving operations. The designed model is useful for planning, tracking and control of earthmoving operations. A Genetic Algorithm is used in conjunction with a set of rules developed to speedup the optimisation process and avoid generating and evaluating
hypothetical and unrealistic crew formations. The model accounts for resources that are available to contractors and it is capable of reconfiguring crew formation dynamically during the construction phase while site operations are in progress. The results have shown that the use of spatial technology (GPS) for data collection provides project teams with a timely, inexpensive, and accurate monitoring tool. In addition, without using of waiting time rule, a GA would have selected unrealistic crew formation although it has the minimum possible cost. Reference [4] presented a methodology for optimising labour and equipment assignment for excavation and earthwork tasks using a Genetic Algorithm. A number of modifications to features of the basic genetic algorithms are conducted to improve the capacity of them. Five GA runs for the labour and equipment assignment problem were performed, each with twenty different solutions. The results showed that the lowest cost solution is achieved in run 1. The other runs all achieved near-optimal solutions with a difference from optimum in a range of 47%. Reference [5] developed a simulation analysis based lean thinking system to improve the performance of drainage operations maintenance activities. Application of lean concepts to drainage operations maintenance crews relied purely on assumptions due to either poor quality data or no data. The integration of lean theory and simulation methodology provides the tools necessary to improve labour productivity for drainage operations maintenance activities. Reference [6] used agent based modelling methods to simulate space congestion on a construction site to explore the impacts of individual interactions on productivity and labour flow. In the agent-based model, each worker and task was represented as an autonomous agent. A simulation for two masonry crews intersecting in space was conducted. Each crew consisted of 3 brick layer agents and 2 helpers. The simulation was intended to study the effects of the sizes of labour crews on productivity. In the current implementation, the skilled workers proceed toward the wall, and complete their nearest tasks. Reference [7] presented a linear programming parametric model formulation for supporting the decisions of construction managers. Exploring the multi-objective nature of decision making in repetitive construction projects has been explored. The project consists of six repetitive units each having six discrete activities repeated each unit. All task dependencies are finish-to-start and each activity is performed by a specific crew. The results showed that strictly scheduling methods like CPM and resource continuity (RSM) can provide answers mainly regarding time dimension but cannot address in an integrated way, cost of work-breaks at each task, penalty costs related to delays in project durations or financial costs arising from inadequate cash flow because of lateness in delivery of partial units. It is shown that MOLPS-LRP mode can address them. In this work, utilisation of resources has not been taken into account. Formation of each crew has not been investigated in details. Reference [8] developed an optimisation methodology which integrates discrete-event simulation with a heuristic algorithm to optimise dynamic resource allocation for construction scheduling. Three kinds of heuristic allocation policies for the crew were considered, “fixed”, “first-use” and the “heuristic” allocation policies. The utilisation rates of each group of crew for different allocation polices are compared. As a result, the optimal allocation leads to the highest utilisation rates of the four groups of crew, the “first-first” policy generates the average once, and “fixed” allocation policy produces the lowest ones. In summary, literature reviewed revealed that there is still a need to develop more crew allocation systems to achieve an optimal allocation of workers in labour-driven industries. The primary objective of this study is to use simulation modelling to analyse effects of resource allocation on labour-intensive facilities performance.

4. SIM_Crew System Allocation Modules

SIM_Crew designed for labour allocation purposes consists of a database, simulation, optimisation, visualisation modules. An interface mechanism is designed to include size of population, crossover and mutation methods, and other entry requirements. See figure 2 for the SIM_Crew modules.

Figure 2: shows SIM_Crew modules
In figure 2, the database module consist of several databases: An excel database is used to provide the simulation module with order specification in each production section. A database using Access is designed to provide GA modules with the crews of workers, processing times, and other labour information. These database are then used to accommodate the outputs of each experiment. A simulation module capable of modelling part of the labour-intensive facility is developed. The modelling part involves process maps and the simulation model itself. The simulation model is developed by converting the developed production process map using special simulation modules. By combining those special simulation modules together in accordance to a specified logic by the developed process maps, the simulation model was then built ready for verification. The Genetic Algorithms module involve the logic of GA’s, other related selection rules, crossover and mutation rules. GA’s codes are written using VBA programming language which is already embeded in ARENA software as a supporting language to add more functionality to the current simulation moduel. Visualisation is the last module where the facility processes is animated in 3D. ARENA 3D player was used to provide the animation.

5. PROCESS MAPPING, SIMULATION AND GENETIC ALGORITHMS

To conduct a structural analysis for the processes involved in the precast manufacturing system, IDEF0 diagrams were used to apply structural methodology and to better understand how to improve manufacturing productivity. The data collection and process modeling was divided into three main phases: Data collection, process modeling, documentation, and the action phases. See figure 3

![Data Collection and Process Modelling Phases](image)

The first phase of modeling processes in figure 3 starts with data collection, essential to ensure sufficient amount of data is collected for modeling purposes. Several techniques were used to collect the required amount of data including: structured interviews, on-site visits and historical document review. The second phase the choice of suitable process-mapping tool. The IDEF family of methods was investigated in detail to determine the most appropriate form. IDEF0 diagrams were chosen to model the functions of system’s processes in terms of graphical representation [9]. It was chosen because IDEF0 can be used to develop a rich process description and facilitate the modeling of the system as a complex system. In the action phase, discrete event simulation methodology was used to simulate the production processes of the labour-driven facility. A precast sleepers manufacturing simulation model was developed and integrated with GA’s for better search capability.
GENETIC ALGORITHMS

The benefit of using Genetic Algorithms in conjunction with a simulation model is to search for an optimum or near optimum allocation of a set of possible crews of workers on each production process. This type of allocation optimises both cost and productivity, under various constraints related to desired production demand, crew availability, and machine condition. The developed GA procedure that suits crew allocation optimisation in simulation models, involves five primary steps, as follows:

STEP 1: DEFINING A SOLUTION REPRESENTATION (CHROMOSOME ENGINEERING)

A solution to the optimisation problem is represented by the set of crews available to be allocated on each production process associated with the problem variables, (number of variables = number of sets of crews available and able to be allocated on each production process)

GAs arrange the solution in a string called a chromosome, having a number of elements (genes) representing the number of variables [10]. A chromosome was designed to resolve this type of problem. Figure 4 shows the designed chromosome for crew allocation purposes.

![Figure 4: Chromosome representation for SIM_Crew problem](image)

In figure 4, each number of each gene would give a crew index number of a set of crew alternatives associated with that gene. i.e, this number gives the index of a crew that would be used in the solution. Each gene has different possible alternatives of crews to be used in the solution. The chromosome is encoded in a decimal way. The chromosome length represents the maximum number of processes involved in the manufacturing system.

STEP 2: SETTING THE VARIABLES, OBJECTIVE FUNCTION, AND CONSTRAINTS

Variables: Variables are the number of sets of crews of workers available to allocate on each production process. (See figure 4 for chromosome structure)

Objective function: To evaluate each chromosome, a single objective function has been adopted to minimise the labour costs. Objective Function = Minimum Cost

Constraints: The user can set limits on production quantity, crew availability limits, and operational hours (shifts).

STEP 3: GENERATING INITIAL POPULATION OF SOLUTIONS (CHROMOSOMES)

As an initial starting point, a population of random solutions (chromosomes) is generated randomly using a Monte-Carlo Sampling Technique, population size and other GA operator’s parameters are influential, affecting the solution and simulation time. In the present model, the user is given the flexibility to input the population size.

STEP 4: EVALUATING THE POPULATION

Once the population is generated, the objective function of each chromosome in this population is evaluated by processing the chromosome into the simulation model, assigning the crew numbers associated with the chromosome of the simulated processes, running the simulation model, obtaining the output costs of labour of that chromosome.
STEP 5: EVOLUTION CYCLES

Simulating natural processes, this initial population of solutions undergoes evolution in cycles called generations. In each cycle, parent chromosomes are selected probabilistically for marriage. Modified operators have been presented to meet the requirements of this type of optimisation problems:

SELECTION PROCESS

In the selection process, all the population size is considered without discarding any of the weak solutions. A randomised selection of chromosomes (Probabilistic Selection Rule) is made for the top minimum costs to be crossed-over and mutated in order to generate a better population. The repetition of the chromosomes is not allowed during any generation so the generated chromosomes should be unique all over the evolution process.

N-POINTS CROSSOVER

The crossover operation in a conventional GA is based on the exchange of genes between two fixed length chromosomes when a bit string representation is used. In the SIM_Crew model, an “n-points crossover” is developed to achieve the random exchanging of genes between each pair of chromosomes. In this type of crossover, n-random numbers will be generated to reflect n places of genes. Then a vertical crossover will be done to swap or alternate subsequently n genes of the first chromosome with the opposite gene of the second selected chromosome. For example, suppose five random numbers have been generated to reflect the places 2, 3, 5, 7, and 9. The chosen places or genes will be exchanged between each pair of chromosomes. See figure 5

M-POINTS MUTATION

The mutation operator is modified to suit the allocation problem being optimised. M numbers of gene will be selected probabilistically and each gene component (crew of workers) are mutated subsequently with the available set of alternatives (available set of crews) for that gene. The places of gene are selected stochastically and the current crew inside that gene is mutated with the available alternatives for that gene. This process is continued until a chromosome unique to be placed into the new population is developed. See figure 6
6. The Case Study

As a test platform for the proposed allocation system, a real-life case study for one of the largest railway sleeper precast concrete manufacturers in the UK was developed. After a detailed investigation of each process in the company, a simulation model was developed using ARENA simulation software to imitate the production processes of concrete sleeper manufacturing. All relevant data was collected by conducting onsite visits, producing and developing flowcharts and process maps including inspection, interviews and stop watch techniques, relevant to data collection requirement. System performance measurements were chosen as useful key performance indicators and included: process idle time, labour costs, production time, and labour utilisation. Relevant decision variables and all other restrictions mentioned earlier in chromosome representation section were considered. Twenty-eight production processes divided between four production lines in two production sections (seven repetitive processes at each) were considered in the simulation model. A set of crews specified to carry out each production process which suggested by the production manager were fed into the system via the database. Probabilistic Strategy was used as the selection strategy, Dynamic N-Points Crossover (N-DPC) and Dynamic N-Points Mutation (N-DPM) were chosen as the crossover and mutation strategies. Fifty generations with a population of size 10 for each were run in order to specify at which generation the results will tend towards stability. In order to analyse improvement opportunities in the sleeper precast concrete manufacturing system, an experimental design study is undertaken in the next chapter to identify the impact of allocation plans on the system performance.

7. Running Experimental Work

The objective of experimentation was to iteratively generate and test the possible crew allocation scenarios until minimum resource allocation cost is found. Genetic Algorithms (GA) module is used to iteratively generate and evolve allocation scenarios and the simulation module is then utilised to test the impact of each allocation scenario on the performance of the manufacturing system in terms of resource utilisation and costs. SIM_Crew, which is based on a genetic algorithm, is proposed for crew allocation with the objective function of minimising total resources cost and maximising utilisation of workers. After running SIM_Crew, results were generated using a population of 10 chromosomes. The chromosomes size being defined to be equal to the number of processes. The number of places to be crossed-over and mutated is defined to be 2, 3 respectively. For the model settings above, a simulation run was performed to check the functionality of the proposed system and determine the type of development required in the model to reduce costs. Next section discusses the results in more detail.

8. Analysis and Interpretation

After running SIM_Crew, a number of key performance indicators were calculated to test the performance of SIM_Crew system. For the resources cost, figure 7, demonstrates the reduction in allocation costs.

![Figure 7: depicts costs reduction using SIM_Crew allocation System](image)

In figure 7, costs were reduced successive through generations, 19% reduction in costs was achieved using the proposed SIM_Crew allocation system. To verify that the GA converges to the optimal solution, the average results were applied to test both improvement and convergence of the solutions see figure 8. SIM_Crew is used to optimise resource utilisation. See figure 9.
This figure shows improvement of the solutions over repetitive generations: generation by generation the average solution is improved due to the evolving process available in the algorithm used. The reduction trend tends to deteriorate during evolution process. Crossover and mutation operators could not explore more than this range for the current Genetic Algorithms settings. Figure 9 shows the optimisation of a worker during the iterative generation, each generation drives worker utilisation in a way to minimise the total allocation cost. Optimised utilisation was synchronized in conjunction with other workers utilisation in order to satisfy the minimum allocation cost goal.

9. Conclusions and Future Work

This paper delivers a GA based simulation model to solve the crew allocation problem. The proposed algorithm is tested using a real life case study and the results are compared with the current operating strategy. The proposed crew allocation concept can successfully provide more flexibility in modeling and could result in solving crew allocation problems in across labour-intensive industries. This flexibility can outperform other methodologies developed by the researchers by: (1) Developing and introducing searching methodologies that resolve this type of allocation problem. (2) Introduces a crew allocation concept that can be used in any labour-intensive industry, especially when repetitive activities are involved. (3) The developed crew allocation system provides more flexibility in terms of calculating idle time, adopting more than working shift and involving a number of priority levels.

A More guided selection rule will be researched to provide a better selection of individuals; A Class Interval (CI) selection strategy will be developed depending on class frequency concept in the descriptive statistics. Dynamic crossover and mutation operators will be improved to add more randomness in the genes exchanging process. A system based on two working shifts will be produced by developing special templates for the process modules to model the real utilisation of resources.

REFERENCES

Optimisation of Operation Sequences in Flexible Manufacturing Cells using Virtual Manufacturing Tools

Magnus Holm *, Josefine Doverborn, Amos Ng & Leo De Vin
Centre of Intelligent Automation
University of Skövde
Box 408, 541 28 Skövde, Sweden

ABSTRACT
Manufacturing organisations are continuously forced to improve the way of working to maintain their competitiveness on the global market. To optimize a production facility requires not only an optimal design of the whole line but also its internal operations sequencing and scheduling during the operational phase. The use of Virtual Manufacturing tools such as Discrete Event Simulation and Computer Aided Robotics has been proven to be highly effective both for production system design and for operational analysis and improvement. This paper proposes a new optimisation method, named SIMBOSeer, which synergistically combines the areas of optimisation, flexibility and virtual manufacturing that integrates robot simulation with simulation-based optimisation. Evaluation of SIMBOSeer, as applied to an existing manufacturing cell at a powertrain manufacturing company in Sweden, has shown that it can be used as an iterative process of analysis and optimisation. The results, when using realistic what-if scenarios, clearly point out that SIMBOSeer can facilitate the optimisation of operation sequences and decrease the total cycle time of the manufacturing cell. This is due to the fact that many non-value adding functions, such as unnecessary tool changes, which have a great negative impact on the effectiveness of the flexible manufacturing cell, can be avoided. Whilst the use of SIMBOSeer has obvious advantages under normal operating conditions of the cell, its use becomes even more beneficial when disturbances like tool failures occur or when unique product variants are introduced to the cell.

1. INTRODUCTION

Today’s industry is forced to be flexible [1, 2] and improve their manufacturing organisation continuously to be able to reach and maintain a leading position on the world market. A support of virtual manufacturing tools in all phases of the production is vital [3, 4] and a strategy that includes the use of these tools for analysis and improvement of the manufacturing operations has shown to be successful. As often before, the car manufacturing industry leads the way in using and implementing such new technologies. Manufacturers like Volvo Cars [5], BMW [6] and Ford [7] have increased their use of these tools. As a consequence of the increasing use, the variety of virtual manufacturing tools is continuously improving and growing, exemplified by robot simulation, simulation based optimisation and multi-objective optimisation. These tools can, when combined with the product development process, provide a foundation for creating a virtual manufacturing environment. The virtual manufacturing environment facilitates concurrent engineering on a world-wide basis, it enables the manufacturers to shorten the lead times not only when introducing a new product but also when changes are implemented in already existing production systems. It is important to use tools that are able to find solutions that avoid interruptions during production and that “provides the production process with the ability to modify itself in the face of uncertainty” [8].

When introducing a product in a flexible manufacturing system three main phases can be defined, namely design, planning and scheduling/control. These three phases need to be considered concurrently for the ability of identifying potential problems in the process at an early stage, which saves time, money and eliminates quality losses. Design and planning is done concurrently to create a representation of the flexible plan. The system’s possible sequences of operations are represented by the flexible plan. The engineer’s possible choices of machines, tools and flexibility are the foundation for scheduling and control [9].

* Corresponding author: E-mail magnus.holm@his.se
As mentioned previously, a manufacturer with an advanced position on the world market must continuously improve the manufacturing system. Often the manufacturer has already optimised the overall production line design. However, further reductions of production time can often be found in the motion between operations and during tool changes. When looking for such improvements it is important to remember that the main aim of an optimisation process must always be to reduce the cost, under constraints such as quality and throughput.

The need of cost control in manufacturing and specifically that the reduction of production time is equal to cost reduction [10], is the aim when optimising the production. Methods for cost reduction are described in [11] which focus on cost pre-calculations and its relations with other manufacturing functions. In [12] it is stated that:

“Determination of the sequence of operations which influences this cost (manufacturing cost) is vital, irrespective of the process or types of products”

The literature review carried out by [13] shows that optimisation within this area has mostly focused on high-level operations equal to the whole workcell, not the low level operations within the workcell. It also shows that very little effort has been spent on the optimisation of the low level sequences of operations within the workcell.

The aim has been to develop a new method that synergistically combines existing methods and available software tools from three research areas, namely: Optimisation, Flexible Manufacturing and Virtual Manufacturing for the optimisation of operation sequences at the Flexible Manufacturing Cell (FMC) level. This new optimisation method is using the traditional variable path design method whilst reducing non-value adding activities such as unnecessary tool changes or unnecessary motions between operations. An example of the latter is a motion which is intended to avoid a potential danger of machine blocking that would not occur anyway.

Due to the short product lifecycle in a flexible manufacturing system, products evolve or new products are introduced which results in changing operations. Changes of the system affect the sequence of operations used to effectively manufacture the product and a new optimal sequence need to be generated. In order to demonstrate that the developed method for optimisation of flexible sequences of operations is able to tackle real-life industrial problems, it has been applied on an existing robot cell. The robot cell is not in actual need of optimisation, but its complexity and number of possible sequences make it appropriate for testing the developed method.

The method to be developed will contain a number of virtual manufacturing tools. The advantages of each tool can be combined to solve the problem at hand. When, for example, joining robot simulation and simulation-based optimisation advantages from both of the methods can be combined. Robot simulation used for continuous simulations of flexible manufacturing cells enable the possibility of a reality based computerised environment where offline programming, kinematics calculations, collision detection and automatic code generation can be carried out. Followed by simulation-based optimisation using genetic algorithms, that can be used to solve complex nonlinear problems with a large solution space. This is especially useful when the number of possible sequences of operations in a flexible manufacturing system easily becomes huge.

2. THE NEW METHOD

This work aims at combining simulation-based optimisation using a genetic algorithm with robot simulation to conduct optimisation of flexible sequences of operations within a flexible manufacturing cell. The goal with the implementation of the method has been to find an optimal sequence of operations, with the least possible of non-value adding functions, so that it minimises the overall cycle time of the cell. Three objectives for the method have been defined to reach the goal of the implementation;

1. It should be possible to use the developed method in both new and existing systems.
2. Use an optimisation algorithm that can recommend a good sequence of operations within a relatively short amount of time.
3. Translate the recommended sequence of operations into the targeted robot controller code.
The developed method can be divided into four main phases as illustrated in Figure 1. To our knowledge no research has been reported that stretches over and combines all of these four phases.

The simulation based optimisation is a combination of two virtual manufacturing tools, one generative and one evaluative. Each virtual tool uses a model of the robot cell. The task for the generative model is to generate candidates of possible solutions or control parameters, in the search for the optimal solution. These candidates are sent to the evaluative model. The evaluative model is executed and the performance measures or the fitness value, using the terms of genetic algorithms, are sent back to the generative model (Figure 2).

The method will be a powerful tool within virtual manufacturing. It supports engineers and operators to find good solutions quickly that adapt to recurring product variant changes and introduction of new products. The name of the method is: SIMBOSeer which stands for Simulation Based Optimisation on low level Sequences of Operations.

SIMBOSeer is a powerful virtual manufacturing tool that foresees a solution for the stated problem. “Simbo” is the name of an island in the Polynesian archipelago. The island is small and consists of three “powerful” volcanoes. A “Seer” is one that can see the future and tell about what is still not known.
3. INDUSTRIAL IMPLEMENTATION

The usability of SIMBOSEer has been tested when applying the method to an existing cell (Figure 3). The cell is not the bottleneck of the production line but its complexity with many operations, flexibility and multiple industrial robots results in an interesting test case for the evaluation of the method.

In the cell the engine blocks are deburred. It is the last cell in its production line, followed by manual inspection and thereafter the engine block is transported to another line for further processing. The elimination is carried out by steel brushes, a total of twelve different tools, with different shapes, diameters and lengths. Twenty deburring operations in total are executed by the robots in the cell. There is no permutation constraints on the sequence in which the operations must be executed so that they can be done in any order. As seen in Table 1, operations are executed on all six sides of the engine block. Due to limitations of the reachability of the robots, a single robot is not able to perform all the 20 operations. The operations on the “front, top and back” of the engine block can be performed by any of the two robots. Operations on the left and bottom side have to be performed by the left robot, robot #1 and operations on the right side by the right robot, robot #2 (Table 1).

The name of the operations is given by the name of the processed holes. Operation T1148 is the name of the operation that deburrs the holes T11 through T48, located on the top of the engine block. During verification of the robot cell, it was discovered that the three operations TP1, FP1 and RVP1 were always executed together and in that specific order. As a consequence of this, the three separate operations, FP1, TP1 and RVP1 were treated by the optimisation as one, inseparable operation. The original sequence of operations and the tools used for the operations are shown in Table 2.
Table 1. Operations of the robotcell

<table>
<thead>
<tr>
<th>Side</th>
<th>Name of operations:</th>
<th>Performed by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>F6162, F1F4 &amp; FP1</td>
<td>Robot #1 or #2</td>
</tr>
<tr>
<td>Top</td>
<td>T1T6, TP1, T1148 &amp; T8182</td>
<td>Robot #1 or #2</td>
</tr>
<tr>
<td>Back</td>
<td>RV6162, RV13 &amp; RVP1</td>
<td>Robot #1 or #2</td>
</tr>
<tr>
<td>Left</td>
<td>RH, RHA, RHB, RHC, RHD, RHP &amp; BRYNE</td>
<td>Robot #1</td>
</tr>
<tr>
<td>Beneath</td>
<td>B8182 &amp; B9196</td>
<td>Robot #1</td>
</tr>
<tr>
<td>Right</td>
<td>LH</td>
<td>Robot #2</td>
</tr>
</tbody>
</table>

Table 2. Original sequence of operations and tools used.

<table>
<thead>
<tr>
<th>Robot #1</th>
<th>Operation</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T8182</td>
<td>#12</td>
</tr>
<tr>
<td></td>
<td>B8182</td>
<td>#12</td>
</tr>
<tr>
<td></td>
<td>RHC</td>
<td>#02</td>
</tr>
<tr>
<td></td>
<td>RH</td>
<td>#08</td>
</tr>
<tr>
<td></td>
<td>BRYNE</td>
<td>#17</td>
</tr>
<tr>
<td></td>
<td>RHP</td>
<td>#14</td>
</tr>
<tr>
<td></td>
<td>RHB</td>
<td>#18</td>
</tr>
<tr>
<td></td>
<td>RHA</td>
<td>#04</td>
</tr>
<tr>
<td></td>
<td>B9196</td>
<td>#13</td>
</tr>
<tr>
<td></td>
<td>RHD</td>
<td>#05</td>
</tr>
<tr>
<td></td>
<td>T1148</td>
<td>#05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Robot #2</th>
<th>Operation</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F6162</td>
<td>#09</td>
</tr>
<tr>
<td></td>
<td>RV6162</td>
<td>#09</td>
</tr>
<tr>
<td></td>
<td>F1F4</td>
<td>#01</td>
</tr>
<tr>
<td></td>
<td>RV13</td>
<td>#01</td>
</tr>
<tr>
<td></td>
<td>T1T6</td>
<td>#03</td>
</tr>
<tr>
<td></td>
<td>TP1</td>
<td>#07</td>
</tr>
<tr>
<td></td>
<td>FP1</td>
<td>#07</td>
</tr>
<tr>
<td></td>
<td>RVP1</td>
<td>#07</td>
</tr>
<tr>
<td></td>
<td>LH</td>
<td>#14</td>
</tr>
</tbody>
</table>

A collision avoidance technique is used by the robots. Zones are declared around and inside the engine block to prevent collision. When executing an operation one or more zones are blocked, the other robot is prevented from working in that or those zones. Since both of the robots perform operations located in the same zone, blocking strongly affects the total cycle time of the cell. The optimal solution is not the same as the shortest cycle time for each robot. An example of a bad solution would be one with a very short cycle time for robot #1 but causing robot #2 to be blocked several times, resulting in a long total cycle time for the whole cell (Figure 4). This individual solution would be rejected by the generative model due to its long total cycle time. The total cycle time of the cell is the fitness value in terms of genetic algorithms, not the cycle time of the individual robots.

![Figure 4. Example of a bad solution](image)

To further evaluate SIMBOSeer, and its capability to find optimal solutions three scenarios were introduced:

1. The existing sequence of operations.
2. A sequence of operations where two existing operations are eliminated.
3. A sequence of operations where one new operation is added.

The first scenario consists of the cell’s existing sequence of operations. In this scenario SIMBOSeer is used to see if the existing sequence is the optimal sequence or if the total cycle time can be reduced. In the second scenario two existing operations are eliminated from the original sequence and in the third an extra operation is added. The added operation can be performed by both of the robots using tool #14. Each scenario used 20 individuals for ten generations in the generative model using genetic algorithms.
4. Results and Analysis

During the original sequence of operations in the first scenario robot #1 is blocked for approximately one minute, robot #2 is not blocked at all. SIMBOSeer needs five generations to find a better solution than the original sequence. When the optimisation process is terminated a solution is found that decreases the original cycle time, 663 seconds, by 14 seconds. The operations T8182 and B8182 use the same tool but they are “separated” in the chosen solution. Even though time is lost when an extra change of tools is performed, the total cycle time is decreased (Figure 5).

To exemplify a simpler version of the original engine, two of the operations were eliminated in the second scenario, B9196 executed by robot #1 and LH executed by robot #2. With the existing sequence of operations kept (without B9196 and LH), the cycle time of the robot cell becomes a little more than 10 minutes, 612 seconds. During the 612 seconds robot #1 is blocked approximately 20 seconds and robot #2 for almost one minute. SIMBOSeer finds a solution that reduces the cycle time by 49 seconds (Figure 5).

There are two major differences in both scenarios one and two between the original sequence of operations and the solution found by SIMBOSeer:

1. No blocking occurs during production.
2. One extra tool change is initiated.

The third scenario with one additional operation represents a more advanced engine than the original in the first scenario. In this scenario SIMBOSeer also have to choose which robot that should execute the additional operation since both robots are able to perform it. The added operation uses tool #14. Placing the added operation last in the sequence is a plausible solution if no optimisation is done during the design phase of the new variant. When placed last in the sequence of robot #1, no blocking of robot #2 occurs since it is already performing its last operation, LH, when robot #1 begins with the extra operation. SIMBOSeer is able to find a solution that reduces the cycle time by 50 seconds (scenario 3:1 Figure 6).
When the additional operation is placed last in the original sequence of operations of robot #2, the total processing time becomes almost twelve minutes, 709 seconds. A possible solution when not using optimisation would again be to place the added operation to the end of the sequence during the design phase. Doing that would not initiate any additional tool change since the operations LH and the added one uses the same tool. Since robot #1 finishes its sequence before robot #2 begins, executing the extra operation does not affect the performance of robot #1. It is still blocked for almost one minute between operations RHD and T1148. A solution with a shorter total production time than the initial sequence is not found (scenario 3:2 Figure 6). Given that the cycle time of robot #1 is much shorter than the cycle time of robot #2, almost one and a half minute even though it is blocked for almost a minute, it does not affect the total production time. Since robot #2 is not blocked and there is no additional tool change that can be eliminated, no optimisation is possible. The additional operation in scenario three should be performed by robot #1.

As shown in the performance graphs (Figure 5 and Figure 6) is SIMBOSeer able to find solutions that significantly reduce the total cycle time of the production cell. The improvements in cycle time can be summarized:

1. Test scenario I Improvement with 2.0 %
2. Test scenario II Improvement with 8.0 %
3. Test scenario III Improvement with 6.9 %

5. CONCLUSIONS AND FUTURE WORK

The aim of this paper was to develop a new method that can synergistically combine existing methods and available software tools from three research areas, namely optimisation, flexibility and virtual manufacturing, for the optimisation of flexible sequences of operations in the flexible manufacturing cell level. After verification and evaluation of the developed method, named SIMBOSeer, using the chosen cell at the powertrain plant, several beneficial results have been shown. SIMBOSeer combines synergistically the areas optimisation, flexibility and virtual manufacturing and optimises operation sequences within the level of flexible manufacturing cells. It can be implemented using available virtual manufacturing software and a successful combination has been the software IGRIP for robot simulation and Plant Simulation for the optimisation using genetic algorithms.
Further interesting findings are that machine blocking had the most negative effect on the total cycle time and is therefore the major non-value adding function to be eliminated by the method. In the literature review [13] were other functions discussed such as transportation and setup time that SIMBOSeer was developed to minimise and if possible eliminate. SIMBOSeer has also proved itself useful to minimise and eliminate other unwanted functions such as machine blocking in the three scenarios. Another interesting finding is that the integration between simulation-based optimisation and robot simulation, to be used as the evaluative model, has shown synergistically results. SIMBOSeer can effectively be used to test different scenarios, which enables the user to evaluate more than one solution to find the optimal sequence of a specific flexible manufacturing cell. Its use becomes even more beneficial when disturbances like tool failure occur or when unique products variants are introduced to the cell. SIMBOSeer are able to design a new sequence of operations and generate robot code, which automatically, can be downloaded into the robots.

Future work will be done in the area of a robot break down where SIMBOSeer is not only able to reschedule the broken robots future operations to the other robots but also relocate the not yet performed operations so that the product affected by the break down still will be processed completely.

REFERENCES

A STEP Compliant System for Manufacturing of Composites in the Aerospace Industry

S. Cantoni1, L. Carrino2, A. Nassehi3, S.T. Newman3, T. Tolio4, A. Valente4* and C. Vitiello1

1Dep Tecnologie e Materiali Avanzati
CIRA, Capua, CE, 81043, Italy

2Dep Ing dei Materiali e Produzione
University of Naples Naples, NA, 80125, Italy

3Dep Mechanical Engineering
University of Bath, Bath, BA2 7AY, UK

4Inst of Ind Technologies and Automation
Nat Res Counic, ITIA – CNR, Milan, MI, 20133, Italy

ABSTRACT

The design of automated manufacturing systems represents a strategic key to global competition, especially for companies that operate in dynamic production contexts. New production technologies and innovative materials as well as the volatile demand for products deeply impact the architecture and behavior of production systems in matching production requirements. In the aerospace sector the huge investment in identifying new product and process solutions is weakly supported by structured frameworks for designing production systems. This paper presents an integrated approach to support the design of production systems operating in the aeronautic context. The study focuses on the design of manufacturing system for the production of parts from composite materials that require composites manufacturing and traditional cutting processes. An Integrated UML Knowledge Platform in which information on product, process and system is jointly modelled is presented. The research concludes by a characterization of new families of manufacturing resources such as prepreg cutting machines in a data model compliant with the STEP-NC standard. A number of case studies will be used to demonstrate the application of the model in real manufacturing scenarios.

1. INTRODUCTION

Companies are recently facing a deep transformation of the global manufacturing. New business models lead to operate in global and networked markets; innovative materials and production technologies support the design of products characterized by complex shapes and advanced performance; digital enterprise technologies allow such companies to concurrently manage different production tasks over the global production network with a meaningful impact on the reduction of lead times and costs.

An impressive example comes from aerospace production in which the advent of composite materials had a huge impact on the role that companies play in that market. On the one hand OEMs, such as Boeing and Airbus [1], transformed their production networks by including the sites in which the composites producers were operating. On the other hand, subcontractors of metal components that already operated in the aerospace supply chain were forced to enhance the manufacturing capability by including new production technologies in order to stay competitive.

In this production scenario the problem of designing production systems which embed different technologies in order to gather the production and machining of composites represents a strategic key. In this paper the production system design phase means to select the set of hardware resources (i.e. CNC machines, carriers and load/unload stations) in terms of type and number in order to match production requirements. The production of composites requires a high level of customization of the system characteristics due to the product characteristics. As a consequence, that traditional architecture of machine tools and auxiliary devices used for traditional aerospace part manufacture do not meet these requirements. In addition, the economical impact of new system configurations in terms of investments and company profit is normally high and, thus, requires detailed planning and evaluation.

* Corresponding author: Tel.: (+39) 340-4690895; Fax: (+44) 1225 386928; E-mail: a.valente@itia.cnr.it
The complexity of the production system design problem is increased by the scarce knowledge of the phenomena involved in the design process, especially in the aerospace context. Over the last few years, a vast body of industrial and academic research has been carried out with the aim of developing new product concepts and more competitive production technologies. The result is, that only a very low percentage of companies operating in this sector have the know-how and the tools to handle the production system configuration problem. This consistently increases the probabilities that machine tool builders sell manufacturing system solutions which do not properly match the customer requirements, i.e. production systems with oversized capabilities in order to increase their profit or old models of machines which are already available in stock and are required to be sold quickly. A further time intensive feature characterizing the production system design task deals with the information exchange between the machine tool builder and the customer [2]. This is mainly due to the absence of a shared vocabulary: in fact the scarce standardization of the information format could lead to a recursive interaction between the machine tool builder and the system user.

The aim of this paper is to provide structured approach to support the design of production systems dedicated to the production and machining of composite components for the aerospace sector. For this purpose, an Integrated Framework for the aerospace, entitled AeroFRAME, is presented.

The first part of the paper provides background on the current technologies for the manufacturing of composites in aerospace and a description of the complexities to be tackled at production system level. The problem statement related to production system design is then introduced, together with the main features of AeroFRAME. The major part of the paper provides a detailed description of the AeroFRAME UML Knowledge Platform which includes a STEP-NC compliant model for the geometric and kinematic characterization of Fabric Cutting Machines. The final sections of the paper shows how the STEP-NC compliant model is applied to different families of fabric cutting machine tools that are available in the market, together with concluding remarks and future developments.

2. RESEARCH BACKGROUND

The aircraft industry represents the leading sector for composite technology applications. The manufacturing of composites traditionally requires a significant manual work. As the use of composite parts on aircrafts is rapidly spreading [3,4], the need to produce components with complex shapes and to increase the production rates represents major challenges to the industry [5,6].

The Hand Lay-up process is the basis for all the modern technologies for composite manufacturing and is still the most diffuse technology. For this reason this work includes a detailed description of the entire process. The four main steps of the Hand Lay-up process which characterize the manufacture of composite components are illustrated in Figure 1. The first step consists of cutting the Prepreg material, e.g. the mix of fabric and resin. This cutting process can be either manual or automated, depending on the productivity requirements and budget constraints (i.e. mechanical cutting with machine tools, non conventional cutting with water jet machines, ultrasonic cutting with ultrasonic cutting machines). The second step is the stratification of plies, Here the number of plies depends on the thickness requirements of final components, the reciprocal orientation of the plies deals with the mechanical requirements of the final components. As with cutting, this process could be manual, semi-automatic by using laser projectors that project the correct orientation of the fibers for each ply or, finally, automated by the means of CNC pick and place anthropomorphic robots. The third step consists of preparing the bag: this step is the only one which can exclusively be manual because of its complexity. It requires the use of different material which are fundamental for the autoclave curing process such as, for example, the use of a breather fabric that absorbs the excess of resin or the peel ply that avoids any adhesion between the prepreg component and the other materials (e.g. with the breather ply and the mould). Once the vacuum bag has been properly positioned and sealed, the vacuum is applied by means of a vacuum pump. The fourth step is the curing process in the autoclave in which a profile of temperature and pressure is set for a time length. The parameters of this process depends on the geometry and the material of the component.

The massive human involvement in this process makes it inefficient for high production demands. Moreover, the quality of final components is sometimes jeopardized by fiber misalignments and/or defects in plies overlapping [6, 7, 8].
In order to improve the process automation and reliability, other technologies have been developed. Automatic Tape Layers - ATL represents an evolution of the traditional Hand Layup process and, currently, is one the most investigated processes for large parts with a average thickness up to 15 cm. A deposition head heats up and compacts the material during the deposition. The head also performs contouring operations contributing to reduce the material scrap [5,9]. ATL is currently used in two basic configurations, a contour tape layer (CTL) for laying tape on a contoured tool surface such as a wing skin, and a flat tape layer (FTL) for laying tape on a flat tool surface. Most ATL systems operating in the world today are open bay gantry style machines with elevated x-axis rails, a cross bridge (y-axis) that moves on the x-axis rails with rack and pinion and/or linear motor drive, and a vertical z-axis mounted on the cross bridge. Tooling (mold) for an ATL part is positioned under the gantry (in the open bay), secured to the floor, the machine is initialized to the tool surface, and prepreg tape “courses” are layered on the tool surface with compaction pressure. Laying of the tape courses is programmable and highly accurate [10]. On the one hand, this process allows to reduce the production time and material wastes by acting on the head path and the overlapping strategy. On the other hand, only single curvature surfaces can be covered by tapes and, at moment, few materials producers are able to provide customized tape widths [8].

Fiber placement - FP process represents a technology that allows to manage narrow tows (3 mm wide) aligned to form a tape. An automatic head adds and drops every single tow during deposition. FP has a fixed tool station (headstock/tailstock) and a machine carriage that moves on an x-axis bed parallel to the fixed tool station. FP delivery head is in contact with the tool (mandrel) surface and lays material onto the tool with low tension and compaction pressure. FP uses multiple individual “tows” or ribbons of slit tape to form a band of material that is placed onto the tool. The FP process has the ability to stop, cut, and restart each individual tow (individual tow control). FP system delivery heads typically process multiples of 12, 24, or 32 individual tows slit tapes to form a band of material. By using multiple individually controlled tows to form the FP process allows to manufacture complex shape and to realize parts with different thickness [5,8,9].

Filament winding is the automated technology characterized by the highest strength-to-weight ratio. It consists of winding resin-impregnated or dried fibers on a mandrel surface in a precise geometric pattern. The mandrel, e.g. the tool, is rotated while a delivery head precisely positions fibers on the mandrel surface [10]. The tows wounded by the filament winding head can be Prepreg or dried [11,12]. By winding continuous strands of fibers, such as carbon fiber or fiberglass, following a predefined pattern, components can be characterized by a strength which is higher then steel at much lighter weights. This technology is mainly applied to the production of pressure vessels and, in general, to axial-symmetric surfaces [11].
Resin Transfer Molding RTM and Resin Infusion Under Flexible Tool RIFT represent two technologies that require the infusion techniques. Dry fibers are deposited in a mould that is put in an autoclave and resin is infiltrated by the use of injection systems. In the autoclave, depending on the materials involved and final shape of components, a proper level of pressure and temperate is applied for the curing process [9,11,12]. Parts produced with these technologies need only few cutting operations after the extraction from the mould and have a good dimension tolerance. Structural part are generally produced by RTM [13] whereas very large parts are produced by RIFT [14,15]. The technologies mainly differ in the use of the tool (closed mold for RTM and open shell for RIFT) and in the resin injection process (pushed by pressure in RTM and driven by vacuum in RIFT).

3. PROBLEM STATEMENT

The production system design problem consists of selecting the best set of resources to satisfy the production requirements during the whole system lifecycle with the minimum expected investment cost. In this work the focus is on Automated Manufacturing Systems (AMS) designed to produce components in composites that require both not-conventional processes (such as for instance Hand Lay-up, Automatic Tape Layers and Filament winding) and traditional cutting processes. For example for Hand Lay-up and cutting processes, the resources which will be taken into account are CNC Prepreg Cutting Machines, CNC Autoclave, CNC Machine Tools, transporters, load/unload stations, pallets and tools.

The design task involves the handle of a wide variety of information ranging from the database of selectable resources (i.e. machines, carriers, load/unload stations, pallets) which can differ in performance, architecture and cost and the data concerning the production requirements, e.g. demand volumes and technological characteristics of the products [16]. Volumes and products features can evolve over time, e.g. products may be modified, new products may be introduced and demand volumes may change. Thus, production information should be also evaluated in an evolutionary perspective since it determines the type and number of resources to be selected over the system lifecycle.

As a consequence, handling the production system design task in an evolutionary perspective in order to match current and future production requirements represents a pretty complex task for companies. This is critical especially for the aerospace production context in which production costs and lead times are very high due to the massive presence of manual operation (both manufacturing and assembly operations) which could be easily automated.

The absence of methodologies and tools which support the decision maker in handling the production system design problem in the aerospace suggests the need of developing a structured framework, titled AeroFRAME, in which a set of fundamental steps are exploited.

3.1 AeroFRAME Framework

The AeroFRAME system design framework proposed in this work is characterized by three main phases as shown in Figure 2.

The first phase consists of collecting input information on present and future production requirements. Input information includes:

- present and forecasted data on demand volume. This information deals with the part mix and the aggregate demand of the family of products which must be produced by the manufacturing system;
- present and forecasted data on geometric and technological characteristics of the part family;
- information concerning the database of physical devices that can be selected to design the manufacturing system, i.e. machines, autoclaves, carriers, load/unload stations, pallets. Data on these resources deal with the architecture (such as number of axes and working cube), performance (such as spindle speed and axes feed) and costs.

The information collected in Phase 1 are used in Phase 2 during the process planning activity. Herein, the setup planning and the pallet configuration are carried out to define a set of alternative process plans for each part type. The results of process planning together with data gathered during Phase 1 are the input of the third phase, i.e. the application of system design models.

As already anticipated, information concerning products and processes can be affected by variability. The ability to forecast future technological and volume changes strongly influences the architecture and flexibility degree of the
designed system. Therefore, the design of manufacturing system lifecycle requires a careful study of present and forecasted production problems over the relevant planning horizon. The production system must be able to react to changes in the production requirements thanks to the characteristics of the system as it is or through reconfigurations. In the latter case, the system must be endowed with the necessary enablers, e.g. modularity and scalability [17].

As illustrated in Figure 2, the skeleton of the whole presented framework is represented by AeroFRAME Knowledge Platform. It allows to store and get information managed within the three main phases of the approach. One of the main difficulties of the system design activity is the management and organization of all the data belonging to different domains which are usually different in their nature. This paper aims at presenting a data formalization framework in which information on product, process and production system is exploited.

![AeroFRAME Design Framework](image)

**Figure 2: AeroFRAME Design Framework**

4. **INTEGRATED UML KNOWLEDGE PLATFORM**

The production system configuration problem involves the management of widely different data related to the product, processes and hardware resources. The development of a common conceptual model in which this information could be framed enables a more effective use of heterogeneous decision support methods at different enterprise levels.

Existing models generally consider products, processes and production systems as separated from each other. Also, the evolutionary aspect which allows the capture of changes in production requirements is generally not addressed. The STEP-NC standard (ISO 14649) aims at integrating product and process data providing a model for data interoperability between CAD/CAM systems and CNC machine tools. However, it does not include information for dealing with the modeling of hardware resources which characterize the production system configuration problem. Many authors presented in the literature object-oriented approaches to model manufacturing systems. Recently a UML-based modeling approach has been proposed in [18] to describe all static and dynamic aspects of a cell controller and a complete FMS. Colledani et al. [19] have developed a formalized link between the production system side and the product/process side in relation to metal components.

The AeroFRAME Knowledge Platform is based on an UML data model consisting of Use Case Diagrams, Activity Diagrams and Class Diagrams that represents an enhancement of the work presented in [19]. The basis of this platform is a data structure in which all the information dealing with product, process and system aspects as well as the reciprocal relation is jointly framed over the production system lifecycle. In this study the concept of evolving information is modeled by scenario nodes and the possibility of manufacturing components in composite materials is taken into account. Figure 3 shows the complete set of classes related to the family of production system information. An example of the detailed information that characterizes the Class Machine, i.e. data on the resource architecture, performance and investment costs is provided in Figure 4.
4.1 STEPMODEL FOR PREPREG CUTTING MACHINES

The selection of physical devices during the production system design requires the ability of evaluating the resource capability and the capability profile over time. The UML representation of physical devices provided in the previous section is too broad to allow a detailed characterization of machines in terms of each single module. Vichare et al. [20] provide a representation of machine tool functionalities that has the advantage of supporting any machine tool functionality and architecture. In their paper they propose a Universal Manufacturing Resource Model (UMRM) which enhances the existing STEP-NC standard (ISO 14649) by including the modelling of CNC machining centre and the auxiliary devices in order to automate the decisions related to the operation assignment and process planning. In the proposed model entities composing of hardware resources, such as mechanical machine elements, kinematic joints and machine axes, are described using the EXPRESS language.

Referring to the Mechanical Machine entity presented in [20] and in accordance with EXPRESS-G representation, the authors propose additional entities in order to allow the representation of resources related to the manufacturing of composite components. In particular a new family of tools have been proposed, titled fabric cutting tools for the cutting of fabrics. The Knowledge Platform introduced in the previous section, highlights the need to extend the resource representation to the modelling of new machine types. In the following section a description of Prepreg Cutting Machine is provided. For the sake of simplicity the additional study has been provided in accordance to the STEP-NC standard. However, the analysis could be easily translated into a UML model in future work. Figure 5 shows an extract of the EXPRESS-G diagram that includes the new entities.
5. APPLICATION CASE STUDIES

The STEP-NC compliant representation for fabric cutting machines consists of framing the mechanical entities and the kinematic joints in order to provide a comprehensive description of the machine functionalities. Figure 6 and Figure 7 illustrate two examples of real two and four axes cutting machines available in the market.

The resource representation refers to a standard legend for representing the various entities related to the fabric cutting machines that is reported in Figure 8.
The geometric models refer to typical two and four axes fabric cutting machine architectures, which have been determined starting from the analysis of two real fabric cutting machines available in the market. The mechanical entities and the kinematic joints have been determined providing a geometric for these machines. For each geometric representation a data model of the instances of the various entities and their reciprocal relationships has been evaluated. Figure 9a and Figure 9b illustrate the geometric and data models for the two axes fabric cutting machines. Similarly, Figure 10a and Figure 10b show the geometric and data models for the four axes machines.

6. CONCLUSIONS AND FUTURE DEVELOPMENTS

The research reported contributes to the current scientific panorama of resource modelling by including the geometric and functional representation of fabric cutting machines for the manufacture of composite materials. The paper demonstrates that a STEP-NC compliant framework can support this new technology area. The standard representation provides a universal data structure during the production system design problem. Future development will be on the extension of the resource modelling to new families of machines for the manufacturing of composites such as a CNC anthropomorphic robot for Filament Winding and CNC machines for Automatic Tape Layers.

REFERENCES

[1] www.species.polimi.it/presentations.html


Monitoring and Control Instrumentation of Gravity Crude Separator

Mohamed Sawessi, Farhad Nabhani, Ian French

School of Science & Technology
University of Teesside
Middlesbrough TS1 3BA, UK

ABSTRACT

Gravity crude oil separators are the standard process for extracting unnecessarily produced water from the liquid stream prior to shipping. To make best use of these separators, it is essential to be aware of their internal performance, and to measure the multi-interface levels between the different materials, such as oil-water, water-solids, oil-emulsion, foam-oil and gas-foam. Thus, the reliability and precision of the level instrumentation, and its ability to monitor all the interface levels of the separator, including the thickness of the foam and the oil-water emulsion, are important when considering the level instrumentation as the main sensing element in the automatic control of the separator vessel. This paper present the different types of oil separator and the current methods and instrumentation used to measure the interface levels, including Electrical Capacitance Tomography.

1. INTRODUCTION

Crude oil as found in nature, requires pre-treatment to improve the product quality and meet the specifications demanded for transportation fuels and related products. These pre-treatment processes include degassing, dewatering and desalination; processes that are complex and require sophisticated equipment and instrumentation. Usually, crude oil from oil wells is pumped to, large tanks to separate gas, oil, water and other materials, such as sand; the ultimate aim being to extract oil. Most crude oil separators rely on gravity separation and the separation process takes a long time [3]. When the hydrocarbon fluid stream enters a three-phase separator, two distinctive phenomena take place. The first phenomenon is fluid dynamic, which is characterized by the gravity separation of oil and water droplets entrained in the aqueous and the oil phases respectively, the gravity separation of gas bubbles entrained in the stream, and the gravity separation of liquid droplets which are dispersed in the gas phase. The second phenomenon is thermodynamic, in the sense that some light hydrocarbons and gas solution flash out of the oil phase and reach a state of equilibrium due to the pressure drop in the separator [4]. If the Separator separates gas from the total liquid stream it is classified as a ‘two-phase’ separator and if it separates the liquid stream into its crude oil and water components it is classified as a ‘three-phase’ separator [1]; Figure 1 [4].

![Schematic representation of a typical separator unit](image)

Figure 1: Schematic representation of a typical separator unit [2]
To optimise the use of a separator, it is important to understand its internal behaviour, to monitor the separation process and to measure the multi-interfaces between different materials, such as gas-foam, foam-oil, oil-emulsion, emulsion-water and water-sand. As far as the thermodynamics aspects of the separator are concerned (i.e., the oil and gas phases). Accurate online measurement of the separator contents is critical for the oilfield operators. Uncontrollable growth of foam and emulsion layers may clog the production system and require an emergency shutdown, while accidental cross entrainment of one phase into the production stream of another (e.g. oil in water) may carry severe environmental penalties. To keep foams and emulsions at bay, expensive chemicals are injected. The inability to measure the actual phase contents leads to their wasteful usage and overdosing can exacerbate rather than cure the problem [4].

2. TYPES OF SEPARATORS

Several different types of crude oil separators are designed and manufactured: horizontal, vertical, spherical, and a variety of other configuration. Each configuration has specific advantages and limitations. However, in this paper we will consider only horizontal and vertical separators.

2.1 FUNCTIONAL SECTION OF A GAS–LIQUID SEPARATOR

Regardless of the size or shape of separator, each gas-liquid separator contains four major sections. Figure 2 shows the common sections in the most crude oil separators.

![Functional sections gas-liquid separator](image)

**Inlet Diverter Section**, because of its high velocity, fluid enters the separator with high momentum. The inlet diverter immediately changes the direction of flow by absorbing the momentum of the liquid, thus, slowing it down and allowing the liquid and gas to separate.

**Liquid Collection Section**, is located at the bottom of the vessel and provides the required retention time necessary for any entrained gas in the liquid to escape to the gravity settling section.

**Gravity Settling Section**, in this section the gas steam velocity drops, and any small liquid droplets entrained in the gas are allowed fall out under gravity to the gas-liquid interface.

**Mist Extractor Section**, gas leaving the gravity settling section will still contain very fine liquid droplets. The mist extractor uses coalescing elements to coalesce and remove these fine droplets of liquid. [5]

2.2 TWO PHASE HORIZONTAL SEPARATORS

The separation of a combined gas-liquid stream into separate phases can be described by the following process. As shown in Figure 3, the combined liquid-gas stream enters the primary separation section (Section A) which contains the inlet nozzle (which is usually tangential) or a diverter battle. This section reduces the velocity of the
fluid stream and takes advantage of the initial effects of a centrifugal or an abrupt change in direction to separate the major portion of the liquid from the combined stream.

![Two Phase Inlet](image)

Fig 3: A schematic diagram of the two-phase gas-liquid separator.

The gas then moves to the secondary section (Section B), where, because of the relatively low velocity, gravity can be used to separate entrained droplets of liquid in the gas stream. The gas stream then moves to the coalescing section (Section C), in which the very small droplets of liquid in the gas stream are removed by impingement on a large surface where they coalesce. Using this arrangement it is possible for the gas stream to leave the separator with less than 0.1 gallon of liquid per million standard cubic feet. The liquid leaves via the liquid collection section (Section D), which acts as receiver for all liquid removed from the mix in the primary, secondary and coalescing sections [6]. The primary operating mechanism in this type of separator is flash separation. Crude oil enters the separator and the entrained light hydrocarbon vapours are flashed. Water is separated from the crude/water mixture because of the difference in densities and the tangential configuration of the stream path, which helps separate water vapour from the crude oil. Thus chemical separation (flashing) and physical separation (centrifugal and gravity) methods occur simultaneously in the separator [6].

### 2.3 Three Phase Horizontal Gravity Separator

Three-phase separators are designed to separate and remove the free water from the mixture of crude oil and water. Figure 4 is a schematic of a three phase horizontal separator. The fluid enters the separator and hits an inlet diverter. This sudden change in momentum does the initial gross separation of liquid and vapour. In most designs, the inlet diverter contains a down comer that directs the liquid flow below the oil/water interface. This forces the inlet mixture of oil and water to mix with the water continuous phase (i.e., aqueous phase) in the bottom of the vessel and rise to the oil/water interface. This process is called “water-washing”; it promotes the coalescence of water droplets which are entrained in the oil continuous phase. The inlet diverter assures that little gas is carried with the liquid and assures that the liquid is not injected above the gas/oil or oil/water interface, which would mix the liquid retained in the vessel and make control of the oil/water interface difficult. Some of the gas flows over the inlet diverter and then horizontally through the gravity settling section above the liquid. As the gas flows through this section, small drops of liquid that were entrained in the gas and not separated by the inlet diverter are separated out by gravity and fall to the gas-liquid interface. Some of the drops are of such a small diameter that they are not easily separated in the gravity settling section. Before the gas leaves the vessel it passes through a coalescing section or mist extractor to coalesce and remove them before the gas leaves the vessel. [7]
2.4 VERTICAL SEPARATORS

The vertical separator is one of the most versatile separators and is capable of handling a wide range of gas/oil ratios. The separator features a fixed vane section installed vertically, perpendicular to flow. As shown in Figure 5 vertical Separators perform four distinct functions – inlet control, vapour demisting, liquid separation, and liquid outlet control. The inlet control section, which contains the inlet diverter, deflects and diffuses the inlet momentum of the inlet fluid and provides bulk gas/liquid separation. The vapour demisting section removes entrained liquids from the gas phase. The mist-laden gas rises up through the vessel, flows horizontally through a distributor plate and the serpentine vanes where the liquid droplets are removed. The gas continues flowing horizontally across the vessel through a second distributor plate to the outlet nozzle. Liquids captured by the serpentine vanes are drained via a down comer to the liquid Separation section. The liquid retention and outlet control section is normally separated from the flowing gas stream by a horizontal degassing baffle. The baffle protects the liquid surface from incoming fluid disturbances and isolates the liquid section from gas turbulence in the inlet section. In this section, liquid is degassed, and in 3-phase designs, oil/water separation occurs [7].

2.5 COMPARISON OF HORIZONTAL AND VERTICAL SEPARATORS

Table 1 gives a brief comparison of the principle attributes associated with horizontal and vertical separators.
3. MEASUREMENT METHODS USE IN SEPARATORS

To make efficient use of the separators, it is important to know what is going on inside, i.e. it is important to monitor the separation process and to measure the levels of multi-interfaces between different materials, such as gas-foam, foam-oil, oil-emulsion, emulsion-water and water-sand. The environment for level sensors varies from near vacuum to very high pressure and from below freezing to high temperature. Because of the wide range of environments the accurate measurement of the multi-interface levels in a crude oil separator is regarded as a challenge for hydrocarbon processing. Many types of level sensor have been developed for this purpose: e.g. externally mounted displacers, differential pressure transmitters, ultrasonic transducers, and single-electrode and multi-electrode capacitance sensors. In practice, however, the need for reliability means that only a limited set of these tend to be employed in the field [9]. As a result, the control strategies currently use in crude oil separators are tend to be quite simple (due to the difficulties in finding suitable multi-interface level measurement systems) and separators tend to be operated somewhat below optimum efficiency.

3.1. EXTERNALLY MOUNTED DISPLACERS.

Externally mounted displacers are commonly used for the interface level measurement [10]. As shown in Figure 6, the method makes use of the difference in densities between two types of fluid being measured, e.g. water and oil. A displacer used for oil/water interface measurement must have a density between the densities of oil and water.
Thus, as the interface moves up and down in a separator so does the location of displacer. Using two different displacers with different densities, both the oil/gas and water/oil interfaces can be measured. However, the presence of an emulsion layer will disturb the measurement and also foam thickness cannot be measured using this technique.

3.2. Differential Pressure Methods

In this method (illustrated in Figure 7) the differential pressure is measured between oil and air, or, oil and water phases. The density of different phases is then used to work out the height of the phases. The primary drawback of this method is that emulsion and foam cannot be measured (and will influence the calculations). Also, the pipe-work connecting the differential pressure transmitters to different phases inside separators is easily blocked by sand and wax formation.

3.3. Ultrasonic Transducers

An ultrasonic level measurement system consists of an ultrasonic transmitter and an ultrasonic receiver, as shown in Figure 8. The returned signals from different interface levels, in the measurement volume, are received by the ultrasonic receiver, as the signals are the function of the density of the medium and the speed of sound in the medium the position of the different interfaces in a separator can be obtained by signal processing [9].
3.4. **GAMMA RAY SENSORS**

In this method, low energy gamma rays are used. A vertical array of low energy gamma-ray sources emits radiation, which is monitored by two vertical arrays of Geiger detectors. The material present in a separator attenuates the radiation ‘seen’ by the detector. The amount of attenuation is determined by the density of the medium and readings are collected by each of the detectors in the vertical assembly. The collected data are then converted to the density measurements. The Geiger detectors can be operated in two modes, current mode and pulse mode. The current mode is used to quantify the intensity of a gamma radiation field and the pulse mode to qualify or analyse the type and energy of radiation. The disadvantage with this type of nuclear interface level sensor is the presence of hazardous radiation [9].

3.4. **A COMPARISON OF LEVEL MEASUREMENT DEVICES**

Table 2 gives a brief comparison of the principle attributes associated with different level measurement techniques.

<table>
<thead>
<tr>
<th>Level measurement technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Externally mounted displacer</td>
<td>• Simple to implement</td>
<td>• The emulsion layer cannot be detected</td>
</tr>
<tr>
<td>Differential pressure transmitter</td>
<td>• The densities of different phases are used for calculation</td>
<td>• Intrusive technique</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Heating and thermal insulations are required</td>
</tr>
<tr>
<td>Ultrasonic transducer</td>
<td>• Can be used for continuous and discrete level monitoring</td>
<td>• It has difficulty in when monitoring emulsion and foam</td>
</tr>
<tr>
<td></td>
<td>• Non-intrusive and easy to install on existing and new separators</td>
<td>• The presence of air bubbles in the liquid attenuates the ultrasound waves</td>
</tr>
<tr>
<td></td>
<td>• Frequent recalibration is not needed</td>
<td></td>
</tr>
<tr>
<td>Gamma ray sensor</td>
<td>• Useful in measuring the vertical distribution</td>
<td>• Presence of hazardous nuclear radiation</td>
</tr>
<tr>
<td>Multi-electrode capacitance Sensors</td>
<td>• Different interface levels can be detected</td>
<td>• Conductive liquids cause problems</td>
</tr>
<tr>
<td>Inductive measurement</td>
<td>• Problems due to conductive liquids can be avoided</td>
<td>• Building up of various insulating layers in a separator</td>
</tr>
<tr>
<td>Thermal measurement</td>
<td>• Can be used effectively for level detection</td>
<td>• A continuous circulating liquid is required</td>
</tr>
</tbody>
</table>

Table 2: A brief comparison of multi-interface level measurement method

4. **NEW TECHNIQUES**

A simple literature review reveals indicates that all conventional sensors, used for the measurement of the interfaces in crude oil separator, are invasive. This is not desirable in industry. Tomography provides a means of measuring the internal behaviour of industrial processes non-intrusively and/or non-invasively. Of various...
tomography techniques available; electrical tomography, which includes electrical capacitance tomography (ECT), electrical resistance tomography (ERT) and electro-magnetic tomography (EMT), was the first to be developed and the most mature. In particular, ECT has the advantages of low cost, high imaging rate, both non-intrusive and non-invasive implementation, no radiation, no moving parts and can withstand HTHP in harsh industrial environment [3].

4.1. Tomography Sensors and Instrumentation

Detection of interface levels between individual layers of media within an oil and gas extraction plant is one of the most basic issues in the petroleum industry. The sensors to perform this task must gather a number of requirements. First, they should be capable to detect all flow components (gas, oil and water - see Figure 11), and their heterogeneous mixtures (i.e. foams and emulsions). Second, they should be safe, inexpensive, and easy to fit and must resist hostile environments inside the plant[11]. Research on industrial process tomography is under development and consists in obtaining estimated images of a cross section of a pipe or vessel containing or carrying the substances of the process. One category of process tomography, is electrical capacitance tomography (ECT), where the cross section to be imaged is surrounded by a set of capacitance electrodes formed from simple metal plates. The vessel wall is made of an insulator, and the electrodes are mounted on the surface, so the measurement is completely non-invasive. In ECT the changes in the electrical capacitance between all possible combinations of electrodes are measured when a dielectric material (for example a mixture of oil and gas) is introduced into the vessel. These changes are caused by variations in the permittivity of that material. From these capacitance measurements, an image based on the variation of the permittivity of the cross section contents can be obtained [11]. The tomography system consists of a capacitance sensor system, a data acquisition system and a model-based reconstruction algorithm implemented on a personal computer. The system can potentially measure the water, oil/water emulsion, oil and foam heights inside a separator on a process plant. A tomography based level measurement device, thus, consists of a set of electrodes placed around the separator (see Figure 9), with an attached data-acquisition and reconstruction unit. The data-acquisition and reconstruction unit, measures the capacitance between the various electrodes, and uses this information to generate a cross sectional image of the phase distribution within the separator. The results can be given as interface heights to a control system and /or a cross sectional image of the phase distribution inside the separator tank [11].

Figure 9: A schematic diagram of a capacitance-based separator tomography unit [11]
4.2. PRINCIPLE OF OPERATION

A schematic diagram of a capacitance-based separator tomography sensor is shown in Figure 9. The Capacitance Sensor Consists of eight electrodes (25 cm long) mounted equally around the separator and an earthed screen (to avoid interference from external fields) surrounding the electrodes, as shown in Figure 12. The measurements are done sequentially. The personal computer controls the multiplexer to select a certain electrode pair, and the capacitance detector, based on a charge amplifier, is used to measure the capacitance. The personal computer reads the measured capacitance into the computer memory, and repeats the process until all measurements have been taken [11]. The measurements are then given as input to the reconstruction algorithm, which transforms the measurements to interface levels. The result is presented numerically by means of interface heights and as an image of the cross section of the separator on a computer display.

![Figure 10: A schematic diagram of an ECT system with its main components.](image)

In addition to the capacitance sensor, a basic ECT system is made up with a data acquisition system and a computer for image reconstruction [12]. An impedance analyser based tomography system. Capacitance data are sampled from two ECT sensors, surrounding the separator and the mixer, to the impedance analyser via the multiplexer box (DAQ), and then sent to a host PC. The PC communicates with the impedance analyser and controls the multiplexer box.

![Figure 11: Parameterisation of the interface levels](image)

![Figure 12: Cross-section of ECT sensor](image)

5. SUMMARY

Multi-phase separation, in particular crude oil separation, is an important process in the oil industry. Because three-phase gravity separators form the main processing component in the upstream section of petroleum production, they have a significant economic impact on the products produced. Conventional sensors used to measure the performance of these separators are invasive. This is not desirable in industry. ECT offers an alternative: it is low cost, it has a high imaging rate, it is non-invasive, it has no moving parts and can withstand high...
pressure and temperature in an unkind industrial environment. It is the intention of the authors to construct a laboratory scale separator/ECT test rig to investigate the potential for using the technique to obtain improved separator control.

REFERENCES


Supplier Development Practice: Arising the problems of upstream delivery for a Food Distribution SME in the UK

Alireza Shokri*, Farhad Nabhani
School of Science and Technology, University of Teesside, Teesside, Cleveland, TS1 5JU, UK

ABSTRACT
The paper aims to emphasize on impacts of the supplier development on reducing the defects in supplier quality for a food distribution Small – Medium Enterprise (SME). An empirical study was conducted to measure the performance of the suppliers in three different key performance indicators of the outsourcing and supplier’s performance to arise the existing problems via information exchange, data collection and data analysis. It has been found that supplier development through data and information exchange and better communication by any food distribution SME raises the problems more promptly. This can dramatically change the supplier’s behavior to improve the quality of the supplier’s service and products. It is suggested that more research is required to raise other key performance indicators and their related problems and to develop more improvement practices. Six Sigma Methodologies could be the potential good practices to be focused in future research studies. Supplier performance measurement, which encompasses data exchange and data collection, develops the systematic flow of information which potentially improves the flow of goods and the whole food supply chain to address the final consumer satisfaction. The research took a novel approach in adopting some transport related key performance indicators of the food supply to the Food distribution and retailing sector which is almost a new approach in food industry.

1- Introduction

Supply Chain Management (SCM) has always been a key element of a successful and competitive business. It consists of managing different levels of internal or external organizational transactions which generally represent the flow of goods and flow of information. There have been different understandings of SCM most of which reflect the need for the customer development and logistics. Arguably, in its most advanced form, SCM is not a subset of logistics but is a broad strategy which cuts across business processes both within the firm and through the channels required to reach the customer and involves the firm’s suppliers[1]. It appears that supplier development alongside the customer development builds the structure of Supply Chain Management. There have been different studies which indicate the proper management of supplier relationships constitutes one essential element on supply chain success [1-4]. The literature has emphasized greater collaboration between the firms and their suppliers to assure an efficient and successful supply chain [5].

Over the last decade, researchers have empirically investigated a variety of research issues that are related to supplier development activities. These issues include critical factors of supplier development [6][7]; the process of supplier development [8-10]; the factors that influence buying firm’s involvement in developing their suppliers[11]; and the effect of technical support provided to suppliers on the performance of both suppliers [12][13] and buyers [14]. Supplier development, as an important construction block of the supply chain, has been adopted in this study to evaluate its effectiveness in a tailored industry with a variety of supply concerns.

This paper intends to focus on supplier development in order to investigate the effect of a systematic quality improvement strategy on improving the supplier’s efficiency and supplier service quality in a food chain. One of the most common characteristics of the food chain is meeting the customer’s requirement, which is challenging. This has been addressed in some literature [8-10] in which supplier development has been introduced as the effective way that improves supplier’s capabilities to meet buyer’s requirement. The most useful elements of supplier development which potentially can improve the supplier’s service quality are examined in this paper.

2- Supplier Development

Supplier Development was pioneered in the automotive industry such as Toyota and Honda which are the masters at supplier development initiatives [15]. The relationship of supplier development practices with performance has been addressed in several studies [3][14][16].

* Corresponding Author: Tel: (+44)1642252121, Fax: (+44) 1642252821, E-mail: Shahab@dlsne.co.uk
The concept of supplier development has received considerable attention from researchers [3] [15] [17-20]. The obvious understanding of supplier development from these studies is the different approaches to defining supplier development.

Wagner (2006) [19] defined supplier development as supporting the supplier in enhancing the performance of their products and services or improving the supplier’s capabilities [19]. Many researchers determined supplier development as an activity which encompasses a long term cooperative effort by buyer firm with its supplier to increase the performance efficiency and/or capabilities of the supplier [6][19-21].

Supplier development is considered as improving the flow of information from the buyer in order to upgrade the suppliers of different tiers. More over, many studies [4] considered supplier development as the assisting activities to improve supplier’s operations. These activities involved:

- Supplier evaluation [14] [18]
- Award Certification [14] [22]
- Providing training and technical assistance [22] [23]
- Establishing effective communication between parties [22] [24-26]

There are also some other researchers that defined supplier development through different dimensions. For instance, supplier development has been considered as the purchasing management or procurement [3] [27], and it has also been defined as the local supplying in which the number of suppliers are rationalized or/and reduced to the minimum in order to establish the longer – lasting relationship with the supplier [3]. The latter definition is focusing on relationship management and building a trust based relationship with more reputable suppliers, while the former definition of Fung(1999) and Cristobal(2005) covers every individual activity which can improve supplier’s operation [3] [27].

The buyers have a more discerning view of their purchasing relationship in order to make sure that their supplier adds value to the product and the service that they provide. Likewise, the supplier not only must provide a good quality product, but is liable to deliver a service quality which will add value to the product. Service provision is also part of an effective supply chain alongside the flow of material from the supplier. De Toni (1994) suggested that the suppliers must show the adequate amount of technological knowledge/ability and the capacity to take care of R & D and design activities [28]. There fore, the supply chain network needs to be capable of information exchange and communication between customer and supplier. This critical issue has urged more researchers to examine different approaches and elements of supplier development. Supplier development divisions according to its application, context and structure have different approaches from a variety of researchers. Cristobal (2005) categorised supplier development in terms of its application and practices including basic, moderate and advanced supplier development [3]. The result of his study is consistent with some other researchers in this case [14] [29] to promote the idea of complementary effect of these three approaches on performance. The application of these approaches depends on the resources of the buyer and level of relationship between the buyer and supplier. The buyer with limited resources is more prepared to apply basic supplier development practices including supplier evaluation, feedback, supplier selection and supplier awarding [3].

Lo (2006) has analysed supplier development through the buyer’s involvement in supplier’s activities. Lo (2006) has indicated that supplier development can be applied both directly and indirectly by the buyer [4]. Direct involvement including awareness of supplier quality, reporting quality problems, evaluating supplier performance and providing feedback and technical assistance is a more effective component in supplier development [4] [19]. Indirect supplier development activities including evaluation and communication are closely linked to direct supplier development and could be regarded as the enabler of direct supplier development [14] [30-31].

Krause et al (1998, 2000) has also indicated supplier development in two different efforts. The first effort is the reactive approach in which the measures are indicated in case of existing poor supplier capability. The second effort is a proactive or strategic approach in which supplier’s performance is improved actively and for the long term, before problems happen [4]. It appears that the latter effort is more challenging, but more effective, since it needs a trust orientated relationship, more supplier’s commitment and more buyers’ resources. The mutual benefits of supplier development in both supplier and buyer have been acknowledged in many literature sources [7][32-33]. The improved supplier’s operation is the most common benefit of supplier development which results in improved product and service to the customer and transmitting the value added product and service to the downstream. This will increase the efficiency of the supply chain. Cannon and Perreault (1999) and Nourdewier et al (1990) have indicated supplier development as the strategically important block to build a strong supplier management structure [19][34-35].
2-1- Benefits of Supplier Development:

Cristobal (2005) and Krause et al (1997b) have focused more on the effect of supplier development on improved operation as the result of supplier development [3][7]. Krause et al (1997b) has suggested through a case study that supplier development could result in a 79% reduction in the number of product defects and a 14% increase in on-time delivery [7]. Supplier development can also potentially be beneficial to the buyer’s firm. As companies search for new opportunities to reduce costs and improve operational efficiency, relationships with their supply base have become a key asset in improving profitability [33].

Wafa (1996) has evaluated the effect of supplier development on JIT through analyzing different hypothesis about the impacts of some supplier development elements on the success of JIT. He concluded that all supplier development elements unanimously have positive correlation with JIT success [32]. Some other literatures directly pinpointed cost reduction for both supplier and buyer via reducing waste (rework and recall), time and paperwork as the result of supplier development practices [21][36-37].

2-2- Elements and Practices of Supplier Development:

Supplier Development practices are the forming elements of building a strong supply management which have been determined as the key successful factors of the supplier development by many literatures [3-4][27]. The key component of effective supplier development is strategic information sharing and long term relationship. The dominant attributes in supplier integration include mutually sharing strategic information and benefit [24-25][38] and establishing long – term buyer – supplier relationship [25-26][38] and mutual trust [24-26]. The level of buyer’s involvement in supplier development practices is an important issue which can categorize the activities in different approaches. Basic supplier development activities include supplier evaluation, supplier selection and supplier awarding [3]. These activities can be practiced either direct or indirect. The level of communication and information sharing could also be different. The literature suggested that the basic communications include the face-to-face meetings, E-mail and Fax, while the advanced communication methods could be Enterprise Resource Planning (ERP) or Electronic Data Interchange (EDI). The use of communication methods broaden and deepen accessible information about business activity and facilitates user participation in a variety of information networks [10][39-41]. The empirical research results also showed that effective information sharing in a timely manner and frequently with supplier is a big step to establish supplier development [10][42].

Different learning activities through mutual information exchange have radical impacts on supplier development efficiency [5]. The learning activities are either single or double loop in which the former is achieved by training methods but the latter is more elusive as it is more challenging to the existing process. [43-44]. It is understood that these elements regardless they are approached direct and indirect or basic and advanced are the key successful factors for an effective supplier development practice. They are required to be adopted adequately and in a mutually trusted environment between supplier and buyer. The communication method is the base of other practices, since the basic communication can affect the profile of information sharing. The activities can also be adopted through a sequential process to have more effective results. For instance, the buyer can evaluate supplier performance, provide feedback, train the supplier and support the supplier to maintain the changes to the operation in accordance to method of communication.

This paper aims to review the problems of supply base within a food distribution SME through identification of key issues, measuring the existing process, finding the root causes of the problem followed by a problem solving case study.

3- Industry Overview

The food service quality supply is the centre of attention in this study. A UK based food distribution with more than £5M annual turnover supplying the food outlets has been selected as a case study, because it meets the features of typical food service SME involving in purchasing management, warehousing, delivery, transport and sales management. The company operates in a tense competitive market and is required not only to meet customer requirements but also to delight the customer by providing the best service and quality product in order to be sustainable in the market. This company is not a manufacturer and its operation is purely providing quality service to the customer with minimum defects. Hence, it is inevitable to transfer the market requirement to the upstream in order to receive the right quality of the service and product. Supplier development could be a promising strategy to establish the value added service and quality in the chain. The company’s suppliers are either UK or EU based and
include the distributors, trading companies, manufacturers and wholesalers.

The base of food and packaging supply met the 80/20 rule, while the 80% of the products were supplied by 20% of the suppliers. The key supplier with more than 50% of supply base was a trading company using the outsourcing. Some other national and international suppliers have also been using the outsourcing in order to deliver the goods to the depot. This had made some complex networks where the company had to communicate with the first tier supplier in order to report the problems about the quality of the product or delivery and await for the response from second tier supplier or manufacturer or outsourcing company via the first tier supplier. Figure 1 represents how the communication between the company as the buyer and the supplier is happening in this market. The figure depicts that the both flow of information and flow of goods are happened directly with the next tier supplier.

Figure 1 – The supplier and customer networking system in this food supply chain

![Diagram](image)

Figure 1 illustrates that S11 as the key supplier received the order information directly, but delivery of the goods is through outsourcing firm. Some other suppliers have also been using outsourcing firms to deliver the goods, although the flow of information from the buyer is communicated direct to the supplier base. Outsourcing features and problems are also key performance indicator (KPI) in problem solving. Outsourcing or third party logistics is generally defined as the provision of a single or multiple logistics services by a vendor on a contractual basis and it has two elements including transport and storage [45]. Exchange of information and measuring the performance of the logistics operation are the key performance indicators for outsourcing [45-46]. Many organisations use outsourcing to maintain their position in market and increase the ability of expanding the market share. Improving of the service, increasing in operational flexibility and reducing the cost could be the possible benefits of outsourcing [45].

The key performance indicators of the supplier development have been selected in order to raise the major problems or defects in supply base. This information was collected via long term operational observation. Delivery condition, on time delivery and product quality were selected KPIs in the selected organization. The associate problems with these key factors were also indicated. Despite of some manufacturing originated problems, sources of the most of problems were logistics and outsourcing. It is pretty important to note that all of the goods were delivered either by the manufacturer owned transport or outsourcing. Moreover, many of goods were stored in the outsourcing storages for a while and this may have dramatic effect on the original quality of the goods. It means, the goods might meet the standards of quality at the point of leaving from manufacturer, but the quality might drop through transport or outsourcing while the goods are in the hand of trading company which requires outsourcing facilities. These supplier associated problems are examined more deeply in the company through this study in order to measure the performance of the supplier and outsourcing.

3-1- Problem 1- There is a poor “Booking In” System by the suppliers:

“Booking In” system is a standard procedure for suppliers in Supply Chain to save the time and effort in transportation of the goods. The company has established the “booking in” system through which, the date and time of delivery by suppliers or outsourcing firms are being recorded. This will help the shop floor team to preplan for the delivery which will result in more streamline operations and faster offloading. The existing “Booking In” system was not effective, since most of the suppliers or third party logistics do not comply with that. The operation of this system was monitored for 13 weeks and the records were analysed in order to assess the supplier performance.
The data analysis revealed that the majority of upstream organizations in this supply chain failed to provide quality service to the company. This analysis was carried out as the part of supplier evaluation which is an important basic element of the supplier development. Table 1 represents that only quarter of the supply deliveries have been booked in which just 12% were delivered on-time.

3-2- Problem 2 - The supplied products are not meeting the quality standard:

The condition of delivery and the quality of delivered goods are two dimensions of the service quality. If the quality of products doesn’t meet the standards, the service quality provided by the supplier will fail. The failure of service quality can have dramatic impacts on business performance, cost, customer satisfaction and profitability of the supplier [47].

The company decided to review the supplier’s performance to assess whether the delivered goods met the quality standards. The deliveries of the suppliers were closely monitored for 24 weeks and the defects were recorded prior to a pilot study. This study identified the number of quality defects associated with supplier’s performance, exposed the potential cause of defects and suggested the action taken by the buyer. The total numbers of 26 supply based defects were recorded for 24 weeks. Every individual defect was treated separately. The result of study is presented by table 2.

Table 2 – The actual supplier’s defects, sources and actions for the Food Distribution organization

<table>
<thead>
<tr>
<th>Defect</th>
<th>Numbers</th>
<th>Possible source</th>
<th>Taken action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reject</td>
</tr>
<tr>
<td>Poor pallet layout in Wagon</td>
<td>4</td>
<td>Logistics</td>
<td>1</td>
</tr>
<tr>
<td>Broken Pallets</td>
<td>3</td>
<td>Logistics</td>
<td>3</td>
</tr>
<tr>
<td>Crushed Packaging</td>
<td>7</td>
<td>Logistics</td>
<td>2</td>
</tr>
<tr>
<td>Poor Wrapping</td>
<td>7</td>
<td>Manufacturing</td>
<td>7</td>
</tr>
<tr>
<td>Contamination</td>
<td>2</td>
<td>Manufacturing</td>
<td>2</td>
</tr>
<tr>
<td>Poor Storage &amp; Handling</td>
<td>2</td>
<td>Manufacturing, Logistics</td>
<td>2</td>
</tr>
<tr>
<td>Poor Labeling</td>
<td>1</td>
<td>Manufacturing</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

This study was conducted to review the key supply problems and assess the buyer’s performance to take action against each individual problem. The study revealed that the company has received the goods in 17 occasions, arranged with the supplier to recall the product in 6 occasions and 3 deliveries were rejected immediately in delivery point. Moreover, the firm as the buyer had effective communication with the supplier through reporting the non-conformances to the supplier, never-the-less; many of the deliveries were received by the buyer. The company was facing with different quality related customer complaints which must be reported to the supplier. There was a recording system available in the company to record, report and monitor the quality related complaints to the supplier. This record was reviewed and the results are presented in table 3. The difference of these data with the table 2 is the source of complaint. The source of the data in table 3 was the customer, whilst the source of the data in table 2 was the company’s quality assessment team.

Table 3 represents 67% of the recorded quality related customer complaints were sourced in supply base. The rest 33% was either unknown or unwanted. Therefore, they could not be classified as the quality related complaints. This part of study revealed that 61% of the supplier related complaints are recalled and reported to the supplier. This indicates that this organization has built an information sharing system with the suppliers in this specific aspect. The complaints have been reported through basic communication methods such as E-Mail, telephone, fax and face to
face visits.

Table 3 – The quality related complaints and the proportion of supplier associated problems

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Numbers of Customer Complaints</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>In which Supplier associated complaints</td>
<td>72</td>
<td>67%</td>
</tr>
<tr>
<td>In which the product has been recalled and reported to the supplier</td>
<td>44</td>
<td>61%</td>
</tr>
</tbody>
</table>

Having identified the key problems, the paper reviewed one of the major problems in its upstream involving one of its 1st tier and 2nd tier suppliers followed by a systematic problem solving procedure to minimize the defects and improve supplier’s performance. It is intended to highlight the company’s role as the buyer in the process of supplier development.

4- Taking corrective action – A pilot case study in Supplier Development

4-1- Purchasing Process Overview:

One of the key products that the company sells is the “Corrugated Pizza Boxes”. This product is bought form the key supplier of this company which is a UK based trading company. It means the supplier takes the order and transfers it to the main manufacturer which is based in the Middle East. The manufacturing firm had no UK market and there was no vision of this market to indicate the UK market standards. So, the quality was not matching with the UK standards. The manufacturing firm produces and loads the Pizza Boxes in the freight containers. Then the product is shipped to the UK market and delivered to this company by the third party logistics. Therefore, there is at least 90 days lead time for this type of delivery. The flow of information including the order sheets, Invoicing and the packing list is conducted through the 1st tier supplier, whilst the goods are delivered direct to the company.

4-2- Problem Overview

The delivered pizza boxes have not met the buyer’s satisfaction due to quality related problems. There was no direct communication with the 2nd tier supplier as the manufacturer used to speak out the problems. The management team decided to take action to reduce the cost of poor quality which are as the following:

- Rejected products
- Reworking
- Inventory and space
- Potential customer loss

As first step, the management team decided to reflect these problems to the manufacturer through the 1st tier supplier. The information was sent to the 1st tier supplier and a high profile meeting was conducted with both suppliers. A steering committee has been established including the management team of the distribution company and the 1st and 2nd tier suppliers to tackle this problem. Then, it was decided that the quality control team in the distribution company to observe, record and measure the data of products quality at delivery point and provide the information to the 1st tier supplier. In fact, this was a huge step in supplier development, since all three parties were closely involved in communication and sharing information. It was also agreed that e-mail, fax and telephone are the best way of communication, as the manufacturing base was in the Middle East and regular close meeting was unlikely to happen.

The quality management team started to monitor the inward containers of the “Pizza Boxes” and record the issues. The measuring criteria for the condition of different areas of product delivery were set up in order to inform the manufacturer about their existing operation. Table 4 represents the quality criteria for different areas which were closely assessed by the quality management team. In fact, this set of criteria would help the manufacturing firm to benchmark their performance and reduce the number of defects. The quality control team assessed the condition of four inward containers coming from the middle–east manufacturer against the measuring criteria. The simple and understandable set of data and quality ranking for each area of the delivery alongside the useful pictures of the defects were provided to the manufacturer firm.
Table 4- The measuring criteria for the quality of different areas in inward delivery of containers of the Pizza Boxes

<table>
<thead>
<tr>
<th></th>
<th>Very Good</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pallet Condition</td>
<td>Impressive</td>
<td>No poor quality</td>
<td>≤ 10% Poor quality</td>
<td>10% - 30% Poor quality</td>
<td>≥ 30% Poor quality</td>
</tr>
<tr>
<td>Wrapping of the Pallet</td>
<td>Impressive tight and Multi layered Wrap</td>
<td>No loose, damaged or poor wrap</td>
<td>≤ 10% loose, damaged or poor wrap</td>
<td>10% - 30% loose, damaged or poor wrap</td>
<td>≥ 30% loose, damaged or poor wrap</td>
</tr>
<tr>
<td>Stacking Condition</td>
<td>Strong, straight &amp; Top level stacking</td>
<td>No poor or leaning Stacking</td>
<td>≤ 10% poor or leaning Stacking</td>
<td>10% - 30% poor or leaning stacking</td>
<td>≥ 30% poor or leaning stacking</td>
</tr>
<tr>
<td>Packaging</td>
<td>Tight, Strong &amp; Perfect Packs</td>
<td>Not tight but No obvious damages or holes on packs</td>
<td>Not tight and very few obvious damages or holes on packs</td>
<td>Not tight and few obvious damages or holes on packs</td>
<td>Not tight and many obvious damages or holes on packs</td>
</tr>
<tr>
<td>Product Quality</td>
<td>Perfect &amp; Impressive quality in every issue</td>
<td>Less than 5 obvious damages on the Pizza Boxes</td>
<td>5-10 cases obvious damages to discarded the products</td>
<td>Few uncommon obvious damages to discard the products</td>
<td>Too many uncommon obvious damages to discard the products</td>
</tr>
</tbody>
</table>

The results of the whole assessment are presented in table 5. It was concluded that the quality aspects were mostly less than average, since many of the aspects were poor. Generally, the condition of the delivery of the Pizza Boxes was not satisfactory, since the condition of 50% “goods in” containers was poor. Figure 2 illustrates the general condition of these containers which exposed the unsatisfactory feedbacks from the food distribution in the position of buyer. As a result, the whole supply chain would be affected by the poor quality of the delivery.

Table 5- Condition of different areas of Inward containers before modification

<table>
<thead>
<tr>
<th></th>
<th>Very Good</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pallet Condition</td>
<td>50%</td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrapping</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stacking</td>
<td>25%</td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packaging</td>
<td>25%</td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Quality</td>
<td>75%</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2- General condition of “inward” containers of the Pizza Boxes before modification

The quality assurance department of the manufacturing firm has decided to look more carefully on the areas where more modification is required. The measuring criteria and the existing performance of the manufacturer were set to enable them to find the root causes of the defect. It was acknowledged by the manufacturer that no other buyers provided such information to help them to improve their performance.

4-3- Taking action Process:

Having analysed the useful information and comments provided by a 2nd tier customer, the manufacturer’s top management team decided to modify their operation to meet the UK market standards which was originally higher than their expectation. The root causes of these defects were identified and the set of following actions was established in production and packaging line to minimize the level of defects:
1. Stricter control all through the production process from the corrugating section to die cutting.
2. Applying two straps to the bundle instead of one.
3. Special supervision at the angel hair removal process.
4. Special supervision at the shrink tunnel process for the bundles.
5. Flat stacking of bundles on pallets. No more stacking on bundle edges.
6. Pallet strapping from four sides.
8. Strict control on in-house forklift drivers (pallet handling) at, production and warehouse / shipping ends.

Apparently, the action number five would affect negatively on the number of items on each pallet and therefore affect on cost efficiency of manufacturer. In order to reduce the problem, it was decided to use different pallet size standards rather than one single pallet size standard. Having implemented these actions on the production and packaging lines, the condition of the delivered freight containers of the Pizza Boxes was analysed for next four inward containers and the result shown in table 5 indicates the significant improvement on the condition of the delivery. The general quality of the condition of each “inward” container has improved and it indicates that the manufacturer firm is achieving to the UK market standard by providing satisfactory delivery to the buyer. Arguably, the root of the problems was identified in production and packaging lines. Therefore, the third party logistics (outsourcing) had little or no effect on the quality of delivery, never the less, the container was in shipment for more than 3 months. Figure 3 illustrates the general condition of delivery after developing the supplier through information exchange.

Table 6- Condition of different areas of Inward containers after modification

<table>
<thead>
<tr>
<th></th>
<th>Very Good</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pallet Condition</td>
<td>30%</td>
<td>70%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrapping</td>
<td>30%</td>
<td>40%</td>
<td>30%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stacking</td>
<td>30%</td>
<td>70%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packaging</td>
<td>70%</td>
<td>30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Quality</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3- General condition of “inward” containers of the Pizza Boxes after modification

Having compared figures 2 and 3, the impact of buyer’s feedback on the performance of the manufacturer is significant, since the proportion of poor aspects in each container was reduced from 50% to 7%. In contrast, the proportion of good and very good aspects of containers has increased from 20% to 80%. The improvement trend on the quality of the containers is depicted in figure 4. It shows that the regular feedback and information from the food distribution company has addressed the continuous improvement of the manufacturer’s performance.

Figure 4- Improvement trend on the quality of supplier’s performance to meet the customer’s criteria in different delivery aspects
5- Conclusion

It is concluded that, supplier development practices in a food supply chain have a dramatic impact on the supplier’s performance in providing better service and product quality for the end consumer. There are different approaches and practices of the supplier development which could be adopted to improve the performance of the supplier. According to the literature review, the attribute of the supplier development for this specific case would be the basic, direct and reactive approach which contains the elements of regular information sharing, feedback, supplier evaluation and basic communications. This type of approach is important in complex supply chain networks such as food supply chain where the flow of information path might be different with the flow of goods path.

6- Future Study

It is recommended to have more research studies on the systematic approach to detect the root causes of the defect in this case. The Six Sigma methodology of DMAIC (Define, Measure, Analyse, Improve, Control) is a reactive method to find the root causes of the defect and optimum solution. The methodology of DMAIC is a reactive approach to develop the supplier, since the Design for Six Sigma (DFSS) is a proactive approach which helps the supplier to improve the performance.

References

Deployment of ERP for Supply Chain Excellence in a Manufacturing Organisation

Stephen Williams, Zeynel Badak, and Chris Thompson

School of Science and Technology
University of Teesside
Middlesbrough, TS1 3BA

ABSTRACT

Supply Chain excellence and increased organisational integration has seen many SMEs adopting Enterprise Resource Planning (ERP) in order to achieve their strategic business aims. Within the current competitive market SMEs are finding the need for manufacturing systems that offer both increased functionality and integration across all areas of the Supply Chain.

ERP is not only seen as an information system but also regarded as a business improvement methodology, therefore ERP offers increased integration in terms of improved communication across all departments with real time information, which potentially can lead to enhanced Supply Chain excellence. ERP offers enterprise-wide integration, which can be linked to effective Supply Chain Management (SCM) in terms of improved internal operations and increased responsiveness to both consumers and suppliers.

The purpose of this paper is to identify the barriers, motives for and the advantages provided from implementing Enterprise Resource Planning within a manufacturing SME in order to provide both increased excellence internally and globally. The use of a unique and dedicated ERP project team within the organisation as a key advantage within the ERP implementation will also be looked at, including the relation of the ERP implementation towards global integration, and the improvements gained for information flow, which was analysed using SCM tools.

The paper will also focus on the utilisation of a Collaborative Innovation Partnership (CIP), which is a scheme funded by One NorthEast, who are Regional Development Agency (RDA) and a primary sponsor for the North East region in the funding of knowledge transfer projects, that have the aim of aiding companies who are looking to develop innovative products, services or processes. The project in this case study involved the deployment of a 6 month project involving three main parties, which included a post graduate, a SME and a University. The project in this case had the aim of aiding both the ERP implementation and the improvement of Supply Chain operations within a manufacturing organisation by using knowledge from both the post graduate and University expertise.

1. INTRODUCTION

Today’s ever increasing competitive manufacturing sector has seen the need for companies to be more responsive in meeting the demands of customers by delivering goods and services that achieve both good quality and value. The current state of the economy has added pressure on small and medium sized enterprises (SMEs), who are finding it more difficult to survive, and therefore are seeking methods to help increase advantage over competitors. In order to achieve increased agility and responsiveness companies are turning to Enterprise Resource Planning (ERP) in order to integrate all information within the organisation, and to help target Supply Chain processes with the aim of improving all internal and external operations.

This paper will focus on the deployment of ERP within a manufacturing SME located in the North East of England. The company has been specialist in the design, manufacture and supply of pipe harnesses, temperature and pressure sensors for more than 30 years and since 2005 has under gone a significant growth due to a demand of pipe harnesses, as a result of a large demand from their main consumer. For the last 10 years the company had been using CMS, which is a legacy system. A Legacy system is an old computer system that is still used by companies because it still provide functionality, despite newer technology on the market [1]. Due to the increase in demand, areas that were ran manually by the legacy system, which included capacity planning and production scheduling, forced the company to look for a more powerful system in order to automate these operations.
The key areas that will be covered in the following sections include a look at the factors that triggered the need for ERP adoption, the implementation process with links to Supply Chain operations and advantages gained from the system including barriers occurred during the implementation. The paper will also look at lessons learned from the ERP process and future actions needed after the implementation of ERP.

2. ENTERPRISE RESOURCE PLANNING ADOPTION

Booz-Allen & Hamilton [2] stated that if an organization, especially a SME are to survive and become more competitive within their market they should make use of improved and enhanced information technologies and enterprise systems, therefore the adoption of ERP is a key consideration in any strategic change. ERP adoption is a key stage because it involves organisations making a decision on whether ERP is right for their needs or what type of ERP system is needed to cover all operations. A report by Klaus [3] identified issues that involved the “lack of research on the adoption of ERP”, which could lead to problems when organisations enter the ERP implementation stage. The adoption process is therefore imperative especially for SMEs who have limited resources.

There are many factors that trigger a company’s need for ERP adoption, according to Parr and Shanks [4], “organisations are mainly looking to develop and enhance flexibility in terms of customer responsiveness and productivity”. Within SMEs reasons behind the adoption of ERP compared to big organisations can be different, factors such as scope, cost and resources of an ERP project will play a big role in the approach taken. Thong [5] identifies two main categories, which includes firstly the “likelihood of ERP adoption” or the process, which leads to ERP adoption and the process in were an enterprise system is selected based on the organisations needs.

2.1. FACTORS WHICH LEAD TO ERP ADOPTION

The first stage involves the reasons organisations adopt ERP. According to Ross and Vitale [6] the main three areas include “technical, operational and strategic”, and are constant with the six motivators, which include improved customer satisfaction, process improvement, reduced operating costs, an enhanced strategic decision-making process, a collective platform and real time data visibility. All these factors will aid in the organisations target to develop and enhance flexibility for customer responsiveness and productivity enhancement.

The manufacturing SME as outlined in the introduction had not only seen a significant growth in the market for their products, but the company had a legacy system that did not perform some of the operations needed for future expansion, which would have resulted in the loss of contracts. These key factors lead to the need for a new information system, which had production scheduling and information integration features.

The company identified the need for Enterprise Resource Planning (ERP) as their main strategic aim in order to automate and integrate information flow and operations both within the UK and Globally. Prior to the ERP selection and implementation the company carried out ERP workshops, which were ran by Teesside Manufacturing Centre (TMC) [7], who work within the University of Teesside and offer manufacturing, design and business process improvement, consultancy and training. The workshops helped to set the scene for ERP business wide implementation.

The company’s worldwide presence has also been a factor that has triggered the need for ERP implementation in order to allow full integration across all organisations, as some subsidiaries act as sub contractors, and therefore the ERP system is seen as a solution that will enable full integration of all organisations across the world using the same system, which will provide integrated information to allow better traceability of production, product development and material flow, and therefore aid Supply Chain excellence.

The ERP adoption for the entire group was approved, with the first implementation carried out in the head organisation in France, the plan for the next ERP implementation was for the manufacturing SME in this case study situated in the UK, with a future plan to adopt the same ERP system in all of the company’s subsidiaries around the world, which include Tunisia, USA, China, India, Korea and Germany.

The company have decided to use a two tier ERP system and according to Davenport [8], an ERP system adoption is “accompanied with the need of considerable investments in terms of time, money and effort” and therefore any organisational decision will require a process which will enable the company to not only evaluate needs in terms of resources, but also allow the company to select the correct ERP system vendor to cover all of the organisations needs. The next page will focus on all the areas that need to be covered in the adoption phase that will enable a successful implementation process.
2.2. ERP Adoption Course of Action

Before an ERP implementation can be undertaken a company must ensure that they are ready for ERP adoption, which is a stage that mainly involves the ERP vendor selection process. The ERP selection process will play a big role in the adoption of ERP, and is undertaken to determine what ERP vendor will offer the required features in order to achieve all objectives and the best service in terms of support.

The development and advancement of new technology has made ERP adoption easier for SMEs, but the biggest development lies with the many different types of ERP systems that can be used and customized to allow organisations to tailor ERP to company needs, therefore reducing the costs on unwanted features and modules. Collegiate Project Services [9] state that the selection of an “ERP Provider is potentially the most important and contentious step in the ERP process”, and therefore organisations want to make sure that the needs for all departments is not only covered but also to ensure that it does not outweigh the organisations resources.

The UK manufacturing SME in this case study decided to use a two tier ERP system, which was selected by all the managers in the relevant departments and by members of the head organisation in France, which allowed the group to select a vendor that covered all the modules needed to achieve their strategic business aims. The selection was based on budget, operational needs and support, which aided in the final selection of the ERP vendor. The selection of the vendor was an important process but there were other areas that need to be covered in order to assess the organisations readiness for ERP adoption.

Figure 1 below outlines the four areas (Organisational Context, Business Processes, Perception of ERP and External forces), which should be exposed in order to be ready for ERP according to Gunasekaran [10]. The manufacturing SME in this case study identified both the external forces in terms of customers, and had a positive perception of ERP. The company also looked at business processes and links with organisation context in order to select the required ERP vendor.

![Figure 1: Readiness for ERP Adoption [10]]

As outlined by Davenport [8] an ERP system adoption is “accompanied with the need of considerable investments in terms of time, money and effort”, and is also linked with the implementation stage, which is a vital stage and discussed further in the next section. Another area that is looked at is the role of ERP and Supply Chain Management (SCM).
3. ENTERPRISE RESOURCE PLANNING IMPLEMENTATION AND SUPPLY CHAIN MANAGEMENT

The ERP implementation stage is seen as the stage that defines an ERP project. According to Martin [11] ERP implementation takes a great effort to get right and has many difficulties, including “the associated organizational change”, which is one of the main reasons that at “least 90 percent of ERP implementations end up late or over budget”. The implementation process can be lengthy depending on the scope, budget, and size of the project and business. Harwood [12] states that an ERP implementation “should not be viewed purely as an IT project” but as an organization change that affects all departments and involves a team effort. Pathak [13] states that there are several suggested methods or models that could be used for ERP Implementation. The methods or stages used in the SME in this case are discussed further below.

3.1. ERP IMPLEMENTATION STAGES

The main stages carried out within the implementation of the ERP system within the manufacturing SME included stages which covered the creation of a strategy plan, ERP team assignment, data cleansing, data migration, user training, creation of procedures, testing, go live and post ERP implementation. Table 1 below provides more details on the different stages used within the ERP implementation process.

<table>
<thead>
<tr>
<th>Implementation Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy Plan and Team Assignment</td>
<td>Development of a project plan and the creation of an ERP Team.</td>
</tr>
<tr>
<td>Data Cleansing and Migration</td>
<td>Cleansing of legacy system data and migration into ERP database.</td>
</tr>
<tr>
<td>Training</td>
<td>Training for users provided by ERP team members and Consultants.</td>
</tr>
<tr>
<td>Operating Procedures</td>
<td>The creation of standard working procedures developed for all users and all departments which used ERP.</td>
</tr>
<tr>
<td>Testing</td>
<td>Testing of all data within a test database by ERP team and consultants before final database was approved.</td>
</tr>
<tr>
<td>Go Live</td>
<td>Legacy system turned off and ERP go live involving all users.</td>
</tr>
<tr>
<td>Post ERP Implementation</td>
<td>Rectify and fix all problems that were not identified or considered during the ERP implementation stage.</td>
</tr>
</tbody>
</table>

Table 1: ERP Implementation Stages

The workshops ran by TMC also helped with the implementation of the ERP as the workshops covered the concept of TIE (Total Integrated Enterprise) [14], which has a “3-step methodology known as CHI and consists of Consensus, Homing-in, and Implementation”. The TIE methodology focuses on the “requirements, organizational culture, business processes, and also understanding the changes brought about by ERP”, which aided the company in planning stages prior and during the implementation process. The KTP associate within the CIP project was also skilled on the CHI and TIE methodology. The training allowed for more focus during the implementation stage, and also helped with the documentation of working procedures and implementation tasks, which is discussed later on within paper. Another essential factor in the implementation stage was the use of an ERP project team within the organisation, and is discussed further below.

3.2. ENTERPRISE RESOURCE PLANNING PROJECT TEAM

One key factor in the implementation of ERP within the manufacturing organisation in this case was the use of a unique and dedicated ERP project team, which was built around two skilled IT professionals who had knowledge of the ERP system as they were involved in the Implementation in France. The ERP Project leader was also the Information Systems Manager, who developed the old system and used SQL Access to transfer data from the old system into the ERP system, which reduced both resource usage and time taken in data migration.

The Team also had the assistance of the head IT Manager in France, who had three years experience using the ERP system, and therefore reduced the need for consultants. The project team also used the assistance of members from Tunisia and USA, which also provided training and development of skills needed for implementation in those countries. Overall the use of a skilled and dedicated project team reduced the costs of resources and consultants, and allowed the project to flow more efficiently. Both Martin [11] and Harwood [12] stated that ERP has an “associated organizational change”, which also affects Supply Chain operations. The organisation therefore needed to analyse the affects of the new ERP implementation. The use of a CIP project to analyse Supply Chain operations and a literature review, which looks at SCM and ERP is discussed further on the following pages.
3.3. SCM AND ERP

The role of Supply Chain Management has helped companies to evaluate and to seek a more effective flow of information and materials. According to Ireland and Webb [15] the use of Supply Chain Management “continues to be adopted by organizations as the medium for creating and sustaining a competitive advantage”. The need of the consumer also plays a large part in the need effective SCM. According to Kaufman [16] the purpose of Supply Chain Management is to “remove communication barriers and eliminate redundancies”, which will aid in the satisfaction of the consumer. The pressure from an every growing competitive market has seen the need for companies to utilize Supply Chain Management tools. According to Mehta [17] the main driving forces behind the use of Supply Chain Management branch from two sources, which are “external pressures and potential benefits from strategic SC alignment”.

Handfield and Nichols [18] agree that Supply Chain Management has external driving forces which result in “increased levels of global competition and creating a more demanding customer and demand driven market”. For an organisation to meet these external and internal pressures they will require Supply Chain tools and technology systems. Tummala et al [19] states that in order for an organisation to adopt Supply Chain Management principles an organisation must have a focus on “all the channels that are employed in the total transformation process from the earth to the end-user to create a productive and reliable Supply Chain network system”. Christopher [20] agrees with this as he states that Supply Chain Management involves the Management of logistics and information flow with suppliers and consumers in order to deliver a better consumer value with an added cost reduction to the full Supply Chain.

The main identified elements that are used in order to achieve an integrated Supply Chain include information systems, which are used for the management of information flow and inventory management using analysis. It is common knowledge that companies need to have a level of information systems for the running of a Supply Chain, and therefore companies are starting to use Enterprise Resource Planning (ERP) systems, which are helping companies to achieve an integrated Supply Chain according to Tummala [19]. The use of Enterprise Resource Planning (ERP) systems helps to provide companies with a solution that will help integrate internal functions and help to reach and provide consumers externally, which will help to add value and improved satisfaction, therefore according to Koh [21] “it is natural that ERP and SCM should be integrated to provide higher business value”. The reason for this is because ERP improves information from all areas such as production, finance, master planning, sales, and human resource management, which aids in providing better Supply Chain focus, according to Wallace and Kremzar [22].

Payne [23] states that the integration of these departments within an organization “enables different divisions and departments to easily share information and communicate with one another”, which will lead to increased Supply Chain efficiency as information is shared and changes can be made quicker. A case study from Bose [24] suggested that the implementation of ERP within Neway showed that the company had vast improvements in Supply Chain operations, which included an increase in outbound order fulfilment, increase in on-time delivery percentage, increased cost savings and better inventory accuracy.

Despite success stories of ERP there are issues that have resulted in implementation failures, which include poor company culture, which according to Beekhuyzen [25] suggests that “Organisational culture plays an important role while implementing ERP systems”, and if a company’s culture is not correct then ERP will fail. Korpela [26] agrees that problems in ERP implementation are results of organisational culture and not technological issues. Other problems with implementation within ERP projects are caused by companies not allowing more employee empowerment, which can lead to poor information being used within the system.

In order to achieve Supply Chain excellence Performance Measurement needs to be carried out to define a benchmark. Performance Measurement has enabled companies to look at data within the Supply Chain and allow a detailed analysis in order to help with SCM functions. Performance Measurement or Key Performance Indicators (KPI’s) help companies to develop measurements and analysis that will include a look at quality, delivery performance, customer satisfaction, productivity and financial performance, which could be used for Supply Chain improvement. ERP is an information system that allows companies to gain data that can be used as a performance baseline [22]. The company felt that Supply Chain operations were important as a part of the ERP implementation and therefore carried out a CIP project, which is discussed on the following page.
3.4. **Collaborative Innovation Partnership (CIP) Utilisation**

The objective of analysing Supply Chain operations with relation to the ERP implementation, and the construction of working procedures involved the utilisation of a Collaborative Innovative Partnership (CIP), which is a scheme funded by One NorthEast [27], and involves the deployment of a 6 month project, which aims to support flexible Knowledge Base-employer partnerships for all employer types, of all sizes, within the North East Region, where maximum benefit from knowledge transfer will arise. The project involved a post graduate who would act as a Knowledge Transfer Associate within the organisation, with the aim of aiding both the ERP implementation and the analysis of Supply Chain operation within the manufacturing organisation. The associate was based within the ERP implementation team in order to carry out a strategic change by reviewing its business processes and the way that it communicates through its Supply Chain.

The project mainly focused on the adoption of Enterprise Resource Planning within the company and its Supply Chain, and assisted the company by aiding in a smoother transition to an ERP environment by having a team member dedicated to the documenting all of the company’s working ERP procedures. The procedures aided all the relevant users within the implementation process by providing not only step by step working instruction but also aiding in the training stage and future training. Other key areas of the project involved the analysis of the existing data within the organisation and its accuracy, which aided in the cleansing and migration of data.

The task most valuable in terms of Supply Chain operations in relation to the ERP was the use of Value Stream Mapping (VSM), which according to Lee and Snyder [28], Value Stream Mapping is a “visualization tool oriented to the Toyota version of Lean Manufacturing” and is used to gain a better understanding of work and data processes throughout a Supply Chain. The results of a VSM will give a company a view of both Value Added activities (VA), Non Value Added activities (NVA) and Essential Non Value Added activities (ENVA), which need more attention in terms of improvement.

The VSM focused on both the material flow and the information flow for a main product within the company. The results with most importance were the results obtained for the information flow, which identified the activities done using the legacy system operations and manual information such as documentation and other paper work. Figure 2 below shows the results that were obtained for material flow, electronic information and manual information times. The results showed that the information process time had the biggest need for improvement, as material operations had good results with a smaller improvement needed. The information processing times indicated that the Legacy system was not efficient enough, which could be improved using ERP.

In terms of information flow the results indicated that it took over 20 hours to process information in order to complete one order, with a manual process time for paper work of over 1 hour. A future map will need to be created in order to determine the new changes and improvements, which will be outlined in the next section including the advantages and barriers from the ERP implementation.

![Figure 2: VA, NVA and ENVA Material and Information Flow Results](image_url)
4. **Enterprise Resource Planning Advantages and Future Actions**

There is a lot of positive evidence for the use of ERP, which can provide many advantages. In terms of barriers or disadvantages ERP has a large failure rate, which according to Lewis [29] “has been estimated at between 66 and 70 per cent”, but on the other hand a study by Koh and Simpson [30] found that not only could ERP “improve responsiveness and agility to change” but also allow “SMEs to create a competitive advantage”, due to an increased responsiveness to change.

The ERP project is still in its early stages after go live, within the manufacturing company in this case and therefore has only had a few advantages. The biggest advantage gained from the ERP implementation was the elimination of manual processes such as capacity planning and production planning, which was replaced by Material Resource planning (MRP), which runs automatically every night and produced both the required planned works orders and planned purchase orders needed to produce and replenish stock. Overall the automated processes had considerable time savings when compared to the VSM map results.

The initial results after full implementation showed a reduction of electronic information processing time of over 5 hours. Another advantage was the reduction of paper work or manual information on the shop floor as the operators will be able to access information from computers placed on the shop floor after implementation, which also seen a reduction in time for manual distribution time. Both improvements are shown below in figure 3.

![Figure 3: Improved Information Flow Results](image)

The analysis of Supply Chain operations did indicate a lack of Benchmarks and other standard measures, which will need to be looked at in the future. The next section will look at the lessons learned from the implementation of ERP and future actions that were identified by the company to enhance Supply Chain operations and the ERP system and methodology.

4.1. **Lessons Learned and Future Actions**

The company did face a diminutive amount of barriers throughout the ERP implementation process, which involved a resistance to change from some users, and there were a few problems caused by the entering of wrong information, which produced errors in the data, and proved to be a lesson learned for all users due to the time taken to correct the errors. Overall the implementation was a success and the problems were eliminated.

An additional problem that was faced by the company was the lack of coverage on some of the system operations when people were on holiday, which could have been resolved if there was a coverage plan in place. Another lesson that was learned was that the original training was not comprehensive enough and should have more detailed to user requirements. The development of procedures did help to cover some of the problems but more detailed training was needed, and has been identified as an essential future action.
The CIP project established that there were no Benchmarks within the company for Supply Chain operations, which was identified as a future action that needed to be performed in line with the ERP system, but only after all functions were running smoothly and all users were content with their daily tasks. Performance Measurement or Key Performance Indicators (KPIs) will help the company to develop measurements and analysis that will look at “quality, delivery performance, customer satisfaction, productivity and financial performance”, which could be used for both Supply Chain improvement and to enhance the company’s Quantity Model [31]. The ERP system has modules that will allow data to be extracted to create new KPI measurements, which will enable the company to establish an excellence standard to measure against.

Felix [32] claims that “You cannot manage what you cannot measure”, and therefore Benchmarking and Key Performance Indicators are a key future action for the company, and will enable feedback and information that will facilitate better Supply Chain analysis and improvement. Camp [33] claims that “Benchmarking is an integral part of the planning and ongoing review process to ensure a focus on the external environment and to strengthen the use of factual information in developing plans”. Benchmarking is also used to help improve performance by looking at weaknesses within the Supply Chain and to help compare Performance Measurements against world-class organisations and therefore best practices can be implemented.

5. SUMMARY

This paper looked at the importance of the ERP adoption process before an ERP implementation can be undertaken, which is seen as the stage that defines an ERP project. One of the key factors in the implementation of ERP within the manufacturing organisation in this case was the use of a unique and dedicated ERP project team, which helped with a successful implementation. The use of a CIP project also provided advantages by aiding in a smoother transition to an ERP environment, which involved having a team member dedicated to the documenting of all the company’s working ERP procedures, and the analysis of internal Supply Chain operations using VSM.

The case study also concluded that ERP has an “associated organizational change”, and that ERP as an integrated information system helps to decrease inventory holdings, increase information throughout all departments and aids in the reduction of unnecessary processes, which has aided in the not only Supply Chain improvements but also aided in the identification of future Supply Chain operations, and has been recognized as a future task to establish KPI’s so that the company could set an excellence standard to measure against for future improvement.

ERP has a large failure rate but if organizations want to improve manufacturing responsiveness and agility, within the competitive manufacturing sector then ERP is a risk worth taking, as results have shown in the SME in this case, where process automation and integrated real time information resulted in reduced process times, streamlined information, and improved Supply Chain operations. The overall assessment of ERP is that it has enabled the company to become more efficient and flexible, which will allow the company to become a main competitor within its current market.

REFERENCES


http://www.onenortheast.co.uk/page/ktpcip.cfm


Designing Supply Contract to Coordinate Supplier-Manufacturer Interactions, when Manufacturer’s Leftover Products have Stochastic Demand

Iman Nossohi and Ali Shahandeh Nookabadi
Department of Industrial & Systems Engineering
Isfahan University of Technology
Isfahan, 84156-83111, Iran

Abstract
Under rapidly changing and highly competitive market, manufacturers should develop new products or services with creative and innovative features and pay attention to customer oriented production. Along considering this issue, in this research we have studied how a manufacturer who is trying to achieve customer oriented production systems, should design supply contract with his major supplier.

A coordinating contract based on option mechanism with two parameters, option price and exercise price, is used as supply contract. Since demand is not naturally deterministic, it is assumed that manufacturer's leftover products at the end of selling season will be sold stochastically. In this environment, necessary conditions for coordination and both parties' optimal decisions are obtained. The analysis shows that the manufacturer's optimal decisions can be the way that satisfies coordination necessary conditions. Also the manufacturer and the supplier both can achieve win-win situations and improve their expected profits in comparison with their expected profits before applying the option contract. A numerical example is conducted to illustrate the analysis. It is believed that the outcome of this research will work for different products especially for innovative products.

1. INTRODUCTION & LITERATURE REVIEW

One of today's biggest managerial challenges in supply chain (SC) is to provide coherent behavior among SC members to achieve channel coordination [1]. Among previous researches on SC coordination, SC contracts have been used as useful mechanisms. Different models of contracts mostly include quantity flexibility contracts, return policies, revenue sharing contracts, discounts and option contracts. In this paper coordination issue is studied in a single-period two echelon supply chain between a manufacturer and a supplier, through option contract.

Tsay [2] and Cachon [3] have reviewed supply chain coordination with contract design. Cachon and Lariviere [4, 5] study a model in a game theoretic set-up with options, and discuss the contracts under both forced and voluntary compliance. Barnes-Schuster et al [6] also study the role of options in supply chain coordination in the two compliance regimes and a two period model. Wang [7] develops a single-period two-stage supply contract with bidirectional options by which the buyer can adjust the initial order both upward and downward. Using kind of option contract, Wang [8] develops a model to study channel coordination and risk sharing in a retailer-led supply chain. There are two main differences among this paper and other previous researches. One is that, we have studied how a manufacturer who is trying to attain to customer oriented production, has to design supply contract with his major component supplier in innovative and competitive products market. Also previous studies on option contract policy assume that any leftover product can be salvaged deterministically, but this assumption is often problematic especially for innovative products and fashion industry. So, the other contribution is that since demand is probabilistic in nature, we have considered stochastic demand for manufacturer’s leftover products in the problem.

The remainder of this paper is organized as follows. In Section two, studied problem is described, section three is devoted to analyze the model before and after applying the contract, section four is numerical example, and finally the paper is concluded in section five.

2. PROBLEM DESCRIPTION AND ASSUMPTIONS

1 Corresponding author: Tel.: (0098) 311-3915505; Fax: (0098) 311-3915526; E-mail: ali-nook@cc.iut.ac.ir
Designing Supply Contract to Coordinate Supplier-Manufacturer Interactions, when Manufacturer’s Leftover Products have Stochastic Demand

A two echelon SC including one manufacturer and one supplier, in a single planning period is supposed. The manufacturer provides his required major component (material or product package) from the major component supplier and after processing them converts them to final product and offers them to retailer or customer.

It is assumed that studied product in this research is kind of innovative product which has some features like long production lead time with the supplier, and completely stochastic demand [7]. Because of long production lead time with the supplier, the manufacturer has to order before beginning of the selling season and the supplier has just one manufacturing opportunity. After the supplier receives the manufacturer's order he will produce components according to manufacturer's order quantity. Because of some logical reasons like benefiting of batch production or prevention of installation costs, it is assumed that the manufacturer will process all of the ordered components and prepare them for the selling season. The manufacturer creates innovation on the product after receiving the component from the supplier. In each echelon of the supply chain, surplus components or products have special salvage values, lost sale shortage (as goodwill cost) and holding costs.

There is a challenge between the manufacturer and the supplier. Due to stochastic demand, it is desirable for the manufacturer to order low, and encourage the supplier to produce more than ordered quantity as capacity reservation meanwhile. On the other hand, because of production costs and long production lead-time, to avoid of extra, unused production, the supplier has an incentive to keep his production quantity according to manufacturer's order quantity.

Paying attention to described interactions, in this research we follow to study this problem to see how the manufacturer, using option contract can create situations in which, not only the supplier tends to reserve production (over production), but also to progress system performance in comparison with decentralized performance. Other assumptions are listed bellow:

1- Each unit of final product needs to just one unit of major component.
2- The supplier doesn't have any constraint for production.
3- Reordering costs for the manufacturer are independent of production quantity.
4- Reordering costs for the manufacturer are ignorable.
5- If it is possible for the customer to take his/her order, he/she will wait for it.
6- All of cost and income parameters like holding or shortage cost are given exogenously. Also an estimation of demand distribution which is invertible is in hand.
7- The manufacturer as contract designer is aware of supplier's decision in response to the contract.
8- Both parties will accept the contract only if their expected profits don’t worse off than the ones before contraction.
9- The created damage because of lack of demand and extra production, for the manufacturer is more than the supplier.
10- In the decentralized system (with a price only contract), both parties can come to agreement about supplier's wholesale price so that, each side's expected profit be positive.

Assumption (10) assures that both parties work with each other before contraction. Assumptions (8, 9) are commonplace. About assumption (9), we have to mention that it is something usual about extensive range of products. For instance in computer and electronically products industry, despite manufacturer's processing costs, salvage value of leftover products will be so low that they are nominated as electronically garbage. On the other hand, unlike the manufacturer's extra products, supplier's extra components may be applicable for different purposes because final process and assembles has not operated on them yet. Other assumptions are all for analytical simplicity. Some of other assumptions that state logical relation between problem parameters are presented in part 3.2.1.

3. MODEL ANALYSIS

Before starting the analysis, let's introduce applied notation throughout the paper:

\( i=1,2 \), index representing supplier and manufacturer, respectively, \( h_i \): Unit holding cost, \( v_i \): Unit salvage value, \( c_i \): Unit production cost (including processing or assembling operations), \( g_i \): Unit goodwill cost, \( w \): Supplier's wholesale price, \( p \): Manufacturer's reselling price, \( Q \): Manufacturer's initial order, \( Q_s \): Supplier's production quantity (capacity building), \( Q_{DS} \): Production quantity in decentralized system, \( Q_{CS} \): Production quantity in centralized system, \( X \): Random demand in the selling season, \( f_X(x) \): P.d.f of \( X \), \( F_X(x) \): Cumulative density function of \( X \), \( Y \): Random variable representing demand for leftover products, \( g_Y(y) \): P.d.f of \( Y \), \( G_Y(y) \): Cumulative density function of \( X \), \( G(y) = 1 - G_y(y) \).

3.1. MODEL ANALYSIS IN DECENTRALIZED SYSTEM
Designing Supply Contract to Coordinate Supplier-Manufacturer Interactions, when Manufacturer’s Leftover Products have Stochastic Demand

In this part decisions of both parties in a decentralized system with a price only contract are investigated. Manufacturer's expected profit is:

\[ \Pi_M = p \left[ Q f(x)dx + \int_0^\infty Qg(y)dy \right] - (w + c_z)Q + g_z \left[ (x - Q)f(x)dx \right] \]

\[ - h_z \left[ (Q - x)f(x)dx + \int_0^\infty yg(y)dy + \int_0^\infty (Q - x)g(y)dy \right] \]

(1)

Using first order condition. Optimal production order is obtained from the next equation:

\[ (p + g_z)F_x(Q) + v_1 \int_0^{Q_{DS}} G_{Tf}(Q_{DS} - x)f(x)dx = w + c_z + h_z F_x(Q_{DS}) \]

(2)

In this situation supplier's expected profit is:

\[ \Pi_S = Q_{DS}^*(w - c_z) - g_z \left[ (x - Q_{DS})f(x)dx \right] \]

(3)

According to assumption (9), both equations (1) and (3) are positive. So it is profitable for both parties to produce the product and it's component.

3.2 Model Analysis After Using Contract

Before analyzing supplier and manufacturer's decisions under contract conditions, at first appropriate contract for this problem must be recognized. With regard to described problem and challenge in part two, kind of option contract like the case which was used by Wang [8] is recognized to be appropriate for this problem. The option contract has two parameters: option price "\( o \)" and exercise price "\( e \)". The option price, which in essence is an allowance, is paid to the supplier for each unit of product reserved beyond the manufacturer’s initial order. When the manufacturer does order a second time he needs to pay "\( e \)" for each additional purchased unit and more than initial order \( Q \). Having known the supplier's production decision in response to the contract and initial order, the manufacturer can maneuver \( (o, e, Q) \) to coordinate the production quantity \( Q_S \) so that channel coordination is reached [8]. The sequence of contract events is presented in [8].

3.2.1 Required Relations Among Contract and Problem Parameters

To prevent of infeasible situations in the problem, it is necessary to define some relations among some of the parameters:

\[ h_1 < v_1 \]  \hspace{1cm} (4)  \hspace{1cm} h_2 < v_2 \]  \hspace{1cm} (5)  \hspace{1cm} w > c_z \]  \hspace{1cm} (6)

\[ p > w + c_z > v_2 \]  \hspace{1cm} (7)  \hspace{1cm} c_z + h_2 - v_1 < w + c_z + h_2 - v_2 \]  \hspace{1cm} (8)  \hspace{1cm} e > v_1 \]  \hspace{1cm} (9)

\[ e + o > c_z \]  \hspace{1cm} (10)  \hspace{1cm} v_1 - h_1 + o < c_z \]  \hspace{1cm} (11)  \hspace{1cm} e + c_z + h_2 > v_2 \]  \hspace{1cm} (12)

\[ e, o \geq 0 \]  \hspace{1cm} (13)

Relations (4) and (5) state that holding extra inventory of products is of value for both the supplier and the manufacturer. Relation (6), states that for the manufacturer, process cost is more than salvage value and less than selling price. Relation (7) represents these conditions for the supplier. Relation (8) is mathematical representation of assumption (9). Relation (9) prevents of supplier's tendency to dispose of the superfluous components instead of responding to manufacturer's reorder. Relation (10) represents logical conditions in which the supplier accepts the contract. Relation (11) prevents of supplier's incentive to produce infinite amounts of components. Relation (12) represents that whenever there is no demand, the manufacturer will not buy supplier's surplus components. And finally, relation (13) prevents contract parameters to be negative.

3.2.2 Supplier's Decision Making

In order to find coordination conditions, at first we have to investigate supplier's decision under the contract. The supplier's profit function is:

\[ \Pi_S = wQ + e \left[ \int_0^{Q_S} (x - Q)f(x)dx + \int_0^\infty (Q_S - Q)f(x)dx \right] + o(Q_S - Q) - c_z Q_S \]

\[ + \left[ (v_1 - h_1) \left[ \int_0^{Q_S} (Q_S - Q)f(x)dx + \int_0^\infty (Q_S - x)f(x)dx \right] - g_z \int_0^\infty (x - Q_S)f(x)dx \right] \]

(14)
Hence the supplier's problem is:

\[ \max Q \quad \text{subject to} \quad E[S \leq Q \geq 0] \]

Supplier's optimal decision about production quantity is represented in lemma 1:

**Lemma 1:** Accepting described option contract, supplier's optimal production quantity is:

\[
Q^*_s = \begin{cases} 
Q & \text{if } Q > \frac{e + o + g_1 - c_1}{e + g_1 - v_1 + h_1} \\
F^{-1} \left[ \frac{e + o + g_1 - c_1}{e + g_1 - v_1 + h_1} \right] & \text{otherwise} 
\end{cases}
\]  
(15)

**Proof.** See Appendix 1.

3.2.3 CONDITIONS FOR SYSTEM COORDINATION

As it was mentioned before, in this paper we want to see how a manufacturer who is trying to achieve to customer oriented production has to design supply contracts. Regarding to customer oriented policies, new idea in innovative products should be flourished from the customer side, and the manufacturer without ordering and holding any inventory, just as a connector, after receiving customer demand, supplying prepared major component from the supplier and completing final product, responds to customer demand. Attaining to these situations needs some considerations especially in the field of production like omitting unnecessary operations. According to this ideal case, the supplier has to be so motivated that without receiving any order from the manufacturer, produces components. Also according to centralized control in the system, one decision maker optimizes system performance [1, 8, 9].

Thus expected profit of centralized system is:

\[
E \Pi_{cs} = (p - c_2) \int \frac{xf(x)dx + nQf(x)dx}{Q} - Qc_1 + (v_1 - h_1) \int (Q - x) f(x)dx - g \int (x - Q) f(x)dx
\]  
(16)

In equation (16), Q represents number of components to be produced and \( g = g_1 + g_2 \).

Using first order condition, supplier's optimal production quantity is:

\[
Q^*_cs = F^{-1} \left[ \frac{p + g - c_1 - c_2}{p + g + h_1 - v_1 - c_2} \right]
\]

The manufacturer as contract designer has to adjust contract parameters so that supplier's decision coincides on ideal and coordinated situation in the problem. The next lemma states coordination necessary conditions:

**Lemma 2:** in order to achieve coordination in the system, bellow conditions are necessary:

i. \( Q^* \leq Q^*_{cs} \),

ii. \( o < \text{MIN} \left\{ \frac{p + g_1 + h_1 - v_1}{p + g + h_1 - v_1 - c_2} c_1 + h_1 - v_1, \frac{p + g_2 - c_2 - v_1}{p + g + h_1 - v_1 - c_2} c_1 + h_1 - v_1 \right\} \),

iii. \( e = \frac{p + g + h_1 - v_1 - c_2}{c_1 + h_1 - v_1} o \).

**Proof.** See Appendix 2.

3.2.4 MANUFACTURER'S DECISION MAKING

In investigation of the manufacturer's optimal decision under framework of lemma two, we first consider second and third coordination conditions and then in part 3.3 we will investigate whether satisfaction of the first condition is possible or not. Manufacturer's expected profit is:

\[
E \Pi_{m} = p \left( \int \frac{Q^*_s f(x)dx}{Q} - (w + c_2) \int \frac{Q^*_s f(x)dx}{Q} - (w + c_2) \right) - (e + c_2) \int \frac{Q^*_s f(x)dx}{Q} - (Q^*_s - Q) f(x)dx
\]

\[
- o(Q^*_s - Q) - g_2 \int (x - Q^*_s) f(x)dx - h_1 \int Q - x f(x)dx + v_2 \int f(x)dx
\]

(17)
4. NUMERICAL EXAMPLE

In this part in order to recognize application of established lemmas and properties, a numerical example is inspected. Suppose a manufacturer wants offer an innovative product to a special market. He has to design a supply contract with the supplier of his required component in innovative product. Assume that parameters of the problem are estimated like bellow (each parameter is for each unit of component or product):

\[ X \sim U[0, 100], Y \sim U[0, 50], c_1=40, h_1=5, v_1=35, g_1=15, w=55, c_2=75, h_2=30, v_2=100, g_2=15, p=160. \]
4.1 SOLUTION

According to lemmas one, two and three supplier's optimal production and manufacturer's optimal order will be obtained. These values are presented in figure (1). According to part 3.3, using the contract, supplier will be encouraged for over production. This point is shown in figure (1), too.

Figure 1: Optimal order and production quantities

Figure (2), illustrates efficiency of the contract for improvement of expected profit in the system. In this figure system expected profit in three cases including after applying contract, before applying contract, and the ideal coordinated one is depicted. It is obvious that using contract leads to improvement of system profit, and win-win situation for both parties. This is the case that was established in the second part of lemma four.

Figure 2: comparison of system profit, before and after contract

5. SUMMARY

In this paper we studied, how a manufacturer who is trying to achieve customer oriented production, should design option contract with his component supplier. Along this purpose the supplier has to be encouraged for over production. We found that supplier will accept the risk of over production, regarding to manufacturer's initial order and contract parameters. The assumptions considered in this study will work for different products with one selling season and stochastic demand, especially for innovative products. Also we show how the win-win condition can be obtained, in which both parties can improve their expected profit, by applying option contract.

Modeling the problem for asymmetric information scenarios where either the supplier doesn't have the demand information or where the manufacturer doesn't have information of the supplier's cost structure, is an appropriate future study.

Appendix 1. Proof of lemma one:

After inspection of first order condition, using lagrangian method, we have:

\[
\begin{align*}
\mathcal{E}(Q_s) + \lambda U(Q_s) &= 0, \\
\lambda(Q_s - Q) &= 0
\end{align*}
\]

Solving these equations, supplier's optimal production quantity will be:

If \( \lambda > 0 \), then \( Q^*_s = F_{\lambda}^{-1} \left( \frac{e + o + g_s - c_s}{e + g_s - v + h} \right) \)

If \( \lambda = c_s - o - (e + g_s)(1 - F_{\lambda}(Q)) - (v - h_s)F_{\lambda}(Q) > 0 \) then \( Q_s = Q \)
Designing Supply Contract to Coordinate Supplier-Manufacturer Interactions, when Manufacturer’s Leftover Products have Stochastic Demand

Combining the two results we can derive (12). □

Appendix 2. Proof of lemma two:
With regard to equation (15), in order to prevent of decentralized control situation, manufacturer's order must be less than the breakpoint in equation (15), on the other hand for conversion of \( Q^*_S \) to \( Q^*_CS \), we must have:

\[
F^{-1} \left[ \frac{e + g_1 - c_1}{e + g_1 - v_1 + h_1} \right] = F^{-1} \left[ \frac{p + g - c_2}{p + g - h_1 - v_1 - c_2} \right] \Rightarrow e + g_1 - c_1 = \frac{p + g - c_2}{p + g - h_1 - v_1 - c_2} \]

Simplifying this statement, we'll have the third condition in lemma two. In this situation the breakpoint in equation (15), will be \( Q^*_CS \), so \( Q^* \leq Q^*_CS \). Until here the first and the third conditions of lemma two are obtained.

Also combining relations (9), (10), (11) and (12) in part 3.2.1 the second condition in lemma two is obtained easily. □

Appendix 3. Proof of lemma three:

\[
\frac{\partial E\Pi_W}{\partial Q} = (e + w - (e + c_2 + h_2 + v_2) f_x(Q) - v_2 \int g(y) dy f(x) dx
\]

\[
\frac{\partial^2 E\Pi_W}{\partial Q^2} = -(e + c_1 + v_1 + h_1) f_x(Q) - v_1 \int g(y) dy f(x) dx \frac{\partial^2 E\Pi_W}{\partial Q^2} < 0
\]

Also combining relations (9), (10), (11), (12), and (13) in part 3.2.1, the second condition in lemma two is obtained easily. □

Appendix 4. Proof of property one:

By letting \( o=0 \), \( Q^* \) that satisfies equation (18) in lemma three (after simplifications), converts to \( Q^*_DS \) that satisfies equation (2). Suppose situation in which, selling manufacturer's leftover products is deterministic and certainly, while other assumptions remain constant. In this situation setting \( o=0 \), leads to:

\[
Q^* = Q^*_DS \Rightarrow F^{-1} \left[ \frac{p + g_2 - c_2 - w}{p + g - c_2 - v_2} \right] \Rightarrow e + g_2 - c_2 = \frac{p + g - c_2}{p + g - h_2 - v_2 - c_2}
\]

Since it is obvious that \( Q^*_DS < Q^*_CS \), if we establish \( Q^*_DS < Q^*_CS \), we will have: \( Q^*_DS < Q^*_CS \).

It can be easily established that for \( a,b,c,d>0 \), where \( a < b, d < c \), \( u \) will be always greater than \( v \):

\[
u = \frac{a + c}{b + d} \Rightarrow u = \frac{a}{b} > v
\]

Assuming: \( a = p + g_2 - c_2 - w, b = p + g - c_2 - d, c = c_2 + v_2 + h_2, d = c_1 - v_1 + h_1 \) We will have:

\[
u = \frac{p + g_2 - c_2 - w}{p + g - c_2 - v_2 + h_2} < \frac{p + g_2 - v_2 + h_2}{p + g_2 - c_2 - v_2 + h_2} = u \Rightarrow Q^*_DS < Q^*_CS
\]

Appendix 5, proof of property two:

Assume \( Q^* \equiv Q^*_DS \), according to lemma one and property one, since \( Q^*_DS < Q^*_CS \), the supplier will produce \( Q^*_CS \).

Regarding to objective function in lemma one the supplier tries to maximize his expected profit, so choosing \( Q^*_S = Q^*_CS \) means his expected profit is more than when \( Q^*_S = Q^*_DS \). It can be inspected easily that supplier's expected profit when \( Q^* = Q^*_CS \) equals with the decentralized one. So, when \( Q^* = Q^*_DS \), supplier's expected profit by over production will be more than his profit in the decentralized system and he accepts the contract. □
Appendix 6: proof of lemma four:

i. In this proof we have to establish that at least one value of \( o \) exists, in all of the cases in lemma three, so that for corresponding \( Q \), the supplier not only accepts the contract but also accepts over production. With properties one and two it can be found that by letting \( o = 0 \), and
\[
\frac{DSQ}{Q} = \frac{Q}{Q_{DS}},
\]
this situation occurs.

ii. According to the last part, when \( o = 0 \) and \( \frac{DSQ}{Q} = \frac{Q}{Q_{DS}} \), supplier's expected profit will increase in comparison with his, in the decentralized one. So, if we ascertain that for \( \frac{DSQ}{Q} = \frac{Q}{Q_{DS}} \), the manufacturer's expected profit will remain unchanged or will improve, it will mean that the total system's expected profit will increase. Manufacturer's expected profit after applying the contract can be separated like below:

\[
E \Pi_M = p \int_0^Q xf(x)dx + p \int_0^{Q_{DS}} f(x)dx - \frac{Q_{DS}}{Q} \int_0^{Q_{DS}} xf(x)dx - p \int_0^{Q_{DS}} xf(x)dx
\]

\[
- p \int_0^{Q_{DS}} xf(x)dx + p \int_0^{Q_{DS}} f(x)dx - p \int_0^{Q_{DS}} f(x)dx - p \int_0^{Q_{DS}} f(x)dx - g \int_0^{Q_{DS}} xf(x)dx
\]

\[
+ g \int_0^{Q_{DS}} f(x)dx - g \int_0^{Q_{DS}} f(x)dx - g \int_0^{Q_{DS}} f(x)dx + g \int_0^{Q_{DS}} xf(x)dx
\]

\[
g \int_0^{Q_{DS}} f(x)dx - h \int_0^{Q_{DS}} f(x)dx + g \int_0^{Q_{DS}} f(x)dx + g \int_0^{Q_{DS}} f(x)dx + g \int_0^{Q_{DS}} f(x)dx
\]

By simplifying 1, 5, 6; and 2, 8; and 7,9; and 10,14; and 11,13; and 12,15, we have: \( E \Pi_M = E \Pi_M^D \)

Therefore manufacturer's expected profit will remain unchanged.

In this situation profit sharing mechanisms can be used to create win-win situation. Let \((1 - \gamma), (0 \leq \gamma \leq 1)\) be the supplier's quota of his extra profit, while giving the rest to the manufacturer

\[
E \Pi_M = E \Pi_M^D + \gamma (E \Pi_S - E \Pi_S^D), \quad E \Pi_S = E \Pi_S^D + (1 - \gamma) (E \Pi_S - E \Pi_S^D)
\]

REFERENCES


Analysing turbulences in sequence fixed production systems according to the pearl chain production planning and control system (ppcs)

Danny Szendrei¹*, Katja Unger¹, Tobias Teich¹ & Tim Neumann¹

¹Department of Business Economy
University of Applied Sciences Zwickau
Dr.-Friedrichs-Ring 2a
08056 Zwickau, Germany

ABSTRACT

Dynamic interdependencies of globalisation force companies and supply networks (sn) of series and mass production to adapt their value creating processes to local conditions of production. The harmonisation of sales order flows with all relevant value added processes seems to be a core action in that progress. To apply this superior goal to the automotive industry, the paper describes an approach of implementing the pearl chain ppcs into running processes. Principal component of the system is the sequencing of production orders according to sales orders. These are to be driven stable and sequence fixed over all components of the value creation system following the pull-principle. In reality though, first trials of running a pearl chain sequence turned out to be disturbed by turbulences during the production process. The paper thus focuses on analysis, evaluation and benchmarking of detected technical and organisational turbulence-drivers in production processes. In a first part, literature research regarding car sequencing algorithms will be presented. In further progress we will describe the segmentation of main production processes into actions of crafting. This is enforced by data collection about local and temporal occurrence of turbulences. The basis for the following analysis of the data will be a process database. Subsequently an analysis of production orders with critical contents at detected turbulence-drivers will be performed. Turbulence drivers then will be investigated and evaluated for possible remodelling by function FMEA. This approach implements an increasing of process stability, the limitation of sequencing criteria and avoidance of turbulences. The overall effect of proceeding that way would be a raise in planning and process stability.

1. INTRODUCTION

The opening or enlargement of markets has been discussed in detail in many papers. Under the heading of globalisation, many articles dealing with procurement and selling markets are discussed [8]. The increasing internationalisation of these markets, as well as a more and more multilateral and shallow division of labour results in a spurting development of complexity in supply networks (sn). This development has to be faced successfully in SCM and business strategies. Thus, clear definition and alignment of objectives has to ensure the survival and the competitiveness of international, multilateral sn’s. The operative part logistics has a significant impact in such scenarios. Logistics, especially SCM, is increasingly impacted by multidisciplinary measures [11]. Operations research, system’s theory, engineering or social sciences enrich logistics research and solutions. Its concepts and solutions are developed, discussed and implemented in academic research as well as in entrepreneurial practice.

Current SCM-business strategies focus on:

- increasing flexibility of products and production,
- implementing lean production criteria,
- aligning of the production strictly to customers demands.

¹ Corresponding author: Tel: +49 375 536 0, E-mail: Katja.Unger@fh-zwickau.de
The key task of SCM in this context is to maximise the stability of all relevant processes of procurement, production and distribution to gain further competitiveness [9]. The combination of different of the above mentioned SCM goals create hybrid concepts, such as the pearl chain ppcs.

In this paper we present certain aspects of implementing the pearl chain ppcs into the running processes of an automotive plant. The key aspects of pearl chain as a logistics concept will be introduced. Conceptual core is the sequencing of production orders (po). These are to be driven stable and sequence fixed over all components of the value creation system following the pull-principle. All components and working stations have to be (re-)designed according to lean criteria. Therefore, we describe the production system and its logistic links. Furthermore, we focus on problems of production order sequencing. As the main problem of po-sequences along pearl chains we highlight the perpetuation and monitoring of the initial sequence over the production process. This approach is meant as a continuation of the work of Boysen, Fliedner and Scholl [12], who investigated car sequencing algorithms in general. Additionally they set up remarkable sequencing criteria. We continuously consider aspects regarding the enhancement of process stability by tracking and monitoring an initial sequence. Thus, we present the current situation in the production system and develop an evaluation procedure. The procedure regards aspects of technical and organisational stability of a po-sequence. A possible discussion about investment into machinery or it-infrastructure is not subject to this paper. Continuatively, the authors consider some aspects of technical restrictions of single po’s, that might result in turbulences within the pearl chain. Furthermore, we will analyse the connection between turbulences and the machinery in the production process. Our whole analysis is being done in order to get a knowledge basis for implementing the pearl chain ppcs into running processes.

2. Current Standards

Initial situation for the analysis of pearl chain ppcs is an automotive plant with the following features:

2.1 Description of the Production System

The pearl chain ppcs is to be implemented in a plant for automobiles. In the plant all work on the car body is done, the body then is painted and finally, assembled with all necessary add-on or built-in modules (electronic parts, engine etc.). In the work of Boysen, Fliedner and Scholl [12], a plant with equal characteristics such as:

- series production,
- multiple working stations along transportation belts,
- mixed model production

are described. Different aspects are buffers between certain working stations and turbulences in the po-sequence.

For our analysis however, we additionally include the possibility of short-term triggering of rush orders or special orders. The plant sources most of the outsourced parts and modules by Just-In-Time or Just-In-Sequence (JIT, JIS) delivery from external parts of the sn. Different models (M1, M2) with various options of configuration are produced in terms of continuous flow production. The production, starting with basis body work and ending with a ready to ship vehicle, is illustrated in figure 1:

![Figure 1: Scheme of car production with process segmentation](image_url)
2.2 The Pearl Chain PPCS

Pearl chain is a current SCM-concept of harmonising and synchronising relevant processes of production, procurement and distribution, that all focus on dynamic demand. In detail the processes of product concept, product development and change management; material procurement, warehousing; production; sales and distribution are considered. The processes have to be implemented in terms of a static flow over all members of the sn [11]. A similar approach is to be found in the work of Schulze and Mohr. There, the combination of the same processes is defined as preventive quality management [2]. The objectives are alike.

Core consideration of the ppcs is the setup of a po-sequence. This sequence is planned some days (5 to 6 days) prior to the start of production and then communicated to all involved external members of the sn [7]. From this time of freezing the sequence, no changes of the sequence are possible. This time prior to final production at the OEM’s plant enables and ensures the procurement, production or assembly of parts and modules at the member’s plants. Furthermore, the requested sequence of modules can be set up in processes of material picking, packaging and transportation. An enlarged planning and producing horizon for external suppliers is achieved in that way.

Figure 2: Scheme of Pearl Chain ppcs

On the day of clearance of the po-sequence, all po’s are to be driven stable, i.e. under retention of the planned po-sequence, over all working stations of the production process. All relevant activities of material disposition and warehousing have to organisationally and functionally be adapted to the po-sequence.

2.3 Production Planning and the Parameters of Sequencing

Many articles concerning production order sequencing can be found in operations research literature. Most of the publications deal with modelling and defining algorithms that shall ensure reliable sequencing. Reliability in many articles targets the aspects of due-time completion of processes or due-time delivery of products [5]. In the approaches of Zhou or Baker [5, 6] the urgency of satisfactory (due-time) order completion towards customers is described and emphasised. Customer’s demands concerning delivery dates as features of entrepreneurial competitiveness are taken into scheduling considerations. Thus, the necessity for safe algorithms and satisfactory sequencing has been growing [14]. The algorithms of the researchers have some parameters for operational scheduling of order sequences, such as:

- $D_{\text{total}}$ total processing time,
- $d_j$, processing time for single operation ($j =$ number of working stations)
- $d_s$, setup time for single operation ($s =$ number of setup actions)

in common. The main criteria for optimization is minimizing the total processing time, and, in that sense, total processing costs as well.

**MINIMIZE** [$D_{\text{total}}$]

The total processing time is defined as:

$$D_{\text{total}} = \sum_{j=1}^{\text{m}} d_j + \sum_{s=1}^{\text{n}} d_s$$  \hfill (1)

In some modern approaches an amplification ratio is multiplied to each single processing time as a matter of increasing planning safety [13], i.e. to minimize time-related disturbances through process uncertainty.

$$D_{\text{total}} = \sum_{j=1}^{\text{m}} (d_j \times a_j) + \sum_{s=1}^{\text{n}} d_s$$  \hfill (2)
Tardy jobs, that delay total processing time are included in most algorithms by considering their occurrence in stochastic measuring. The sum of delayed processing time is then added to the total processing time \[6\].

\[
D_{\text{total}} = \sum_{i=1}^{\infty} (d \ast a) + \sum_{j=1}^{\infty} d_j + \sum_{d=1}^{\infty} d_d
\]

(3)

with: \(d_d\) = delay time for single operation

Another common feature of many solutions is the missing tracking of order processing in a production system according to the initial setup. All solutions take into consideration current criteria regarding competitiveness but neglect the effects to production systems. It is assumed that a generated sequence remains stable while processing it. Doing so, technical resources of production have to cope with the stability necessity. The production system in praxis proved too far away from such necessary stability. Thus, we perform an initial analysis of existing, running systems, as well as a diagnosis for future availability. Our methodical approach will be presented in section 4.

2.4 PRODUCTION PLANNING AND PRACTICAL ASPECTS

The generation of po-sequences for a single working day is restricted by certain criteria. All known criteria are summarised in an area \(H_{\text{sequence}} [12]\) with \(H_{\text{sequence}} \rightarrow \{ h_n \mid n=1, \ldots, k \}\) elements. To improve planning conditions in order to avoid po-turbulences, \(H_{\text{sequence}}\) has to be filled with more restricting elements. In dependence on production program planning, the restrictions are used to generate daily po-sequences. Most important restriction is the delivery date of a vehicle. REMARKING THE DELIVERY URGENCY IN SEQUENCING ALGORITHMS IS ALSO A DISPARITY TO BOYSENS ANALYSIS [1]. It is forced by strategically focussing on dynamic sales order flows in the pearl chain concept. Delivery urgency of real customer orders itself is then restricted by short-term triggering of rush or special orders. These orders can include some hundred vehicles, to be delivered to sales agents or distributing centers according to marketing activities. These vehicles (po’s) are not ordered by customers. Furthermore there are quantitative restrictions to material, supplied by external partners. This case might happen through contractual agreement on supply amounts. Another example for sequencing restriction is the setup of paint blocks. In order to minimize expenses for cleaning and adjustment of the paint shop in the flow of different colours, a po-sequence with homogenous basic paints and lacquers is aimed for. Finally, we introduce the problem of triggering delayed orders into a po-sequence. The delay can be caused by various activities of rework on single orders. Such rework disrupts the sequence on a certain day. According to the time-related amount of necessary rework, this order might easily eradicate into the sequence of the following day [11]. Since that turbulence-order has a certain urgency towards completion time itself, it has to be re-triggered as soon as possible. This certainty has not been taken into consideration in the above mentioned sequencing models (see par. 3.3.1). Initial point for process analysis and following optimization is a known and fixed po-sequence, at least after its initial setup. Since the pearl chain sequence is setup a few days prior to order clearance, such turbulences need special treatment.

2.5 TURBULENCE-DRIVERS IN PO-SEQUENCES

As mentioned in preceding paragraphs, the occurrence of turbulences degrades process stability and endangers all competitive and efficiency potentials of the concept. The processes of procurement, disposition and distribution are synchronised to the initial sequence. Once, certain pearls fall of the chain, all effort of material supply or machinery setup becomes disordered as well. To maintain the potentials of the concept, turbulence drivers require a detailed analysis in order of avoidance. Turbulences occur at certain working stations. So far, turbulences were verified by chronologically counting the po’s when arriving and leaving a working station. Thereby, incoming and outgoing sequences have to be equal. Moreover it has to be analysed whether turbulence at working station \(A_1\) causes other turbulences at a downstream working station \(A_{1+n}\). With this we have to consider possible coherencies between the work being done at both stations. Furthermore, it has to be recorded, what certain contents of crafting is attached to the troublesome po, as well as to the preceding and subsequent po’s. Turbulence drivers thus will be categorised in 2 classes:

1. technical reasons:

There is a lack of stability and reliability at the machinery in the production process. The parameters of machinery have to be adjusted to the parameters of production of all possible variance. Since OEM vehicle production is characterised by a great variance [9, 12] a ranking of turbulence driving working stations should be achieved.
2. organisational turbulence-drivers

As mentioned above, the urgency of completion/delivery is the most important sequencing criteria. Regarding this restriction exclusively neglects the complexity over all po’s in combination. We assume some critical combinations of po-contents. Thus detecting such critical contents in predecessor-successor-combination is necessary. These pairings can be recorded and avoided in further scheduling operations. Particularly regarding technical reasons of turbulences, a short survey of maintenance aspects has to be discussed.

2.6 MAINTENANCE OF PRODUCTION SYSTEMS

As described in par. 1, operational safeness of the production system is a crucial part in a pearl chain scenario. Safe and stable process steps together with their cost effectiveness ensure the physical transformation of po’s in their sequence. Heng and Zhou [1, 3] describe operational safeness and systems availability as direct parameters on competitive impact. Improving and sustaining machinery and process safeness are subject to strict alignment of a companies production management and its maintenance concept. Current maintenance concepts include specific conditions of machines into activities of maintenance. This condition-based maintenance (CBM) [1] is based on a detailed data basis. Over a period of time a set of indicators for following instability or machine failures is documented. Furthermore, the stochastic appearance of those indicators is used for future stability prediction (prognostics) [5] When implementing the pearl chain ppcs, these requirements towards stable production systems have to be met. This includes a strict definition of system stability regarding sequence parameters. Continuously changing single machine settings according to po-contents must not harm the sequence. Before maintenance aspects are formulated for the pearl chain objectives, the elements of the production system have to be checked for the pearl chain parameter flexibility. This paper does not focus on procurement and installation costs in plants, but on the necessity to cover a high variance of products with a lined up production belt productively. In other words, flexibility criteria have to be met.

3. ENVIRONMENTAL DESCRIPTION

3.1 FLEXIBILITY

We introduced flexibility as a key factor of competition in par. 1. Flexibility ensures short time reaction to changes in demand on markets [9]. The automotive industry is affected very intensively. This is caused by an increasing variety of models with a huge amount of different configurations. A variance of more than 10,000 different combinations of models and fittings has to be handled reliably and on time. With respect to mass customisation we differentiate between volume flexibility and product mix [9].

Volume flexibility summarises all measures to dynamically and quickly change the amount of output. Of course, all processes of procurement and production are impacted by the output rate. Product mix covers efforts to dynamically change between certain products in production systems. According to Hallgren and Olhager [9], both forms of flexibility are interrelated. Regarding the pearl chain ppcs, flexibility criteria have to be implemented into the production system. On one side, the system has to cope with changing outbound vehicles. Safe and due time distribution has to be ensured. On the other side, all production segments must cover a continuous proceeding of almost unique pos. The working stations must provide for sequence fixed po-processing.

3.2 LEAN PRODUCTION

A further pearl chain consideration is lean production. Lean production comprehends the identification and elimination of all waste. These measures extend to the whole value adding chain [14]. According to Bayou and de Corvin, lean production lacks a strict definition [8]. Thus, all conducted measures to reduce production processes to performance-relevant actions at different plants are incomparable. Our approach does not intend to install a comparable benchmark system in terms of leanness. By evaluating the processes according to our FMEA procedure though (see par. 4.2), we will be able to identify unessential processes, i.e. to highlight rework. The evaluation of rework and its reasons can lead to instable processes. Process redesign in order of avoidance cuts such processes.
4. METHODOLOGY

4.1 PROCESS DATABASE

The methodical approach of sequence analysis will be accomplished on the basis of a process database. As current solutions for maintenance of production systems are set upon empirical and statistical data it should be possible to use such an information platform for process diagnostics as well. Regarding the productive transformation of po’s in a flow system we can describe the whole processing system as subject to stability and reliability measures. Subject to the analysis are working stations and their causing of turbulences. In that sense, information about the

- po-sequence,
- po-contents and po-identification,
- troublesome working stations

need to be collected. All necessary information concerning those elements have to be covered in a database structure. This structure includes the following entities: po-ID, po-content, vehicle class, po-process history, production date, production shift, supervisor, scanning points and detected defects. With that structure, the po-sequence of any period in reality can be verified with the planned sequence. Furthermore, the sequence of the same po’s, being processed in a production subsystem AS$_{1+n}$ can be verified against the sequence in AS$_1$. Optimal case would be two complete identical sequences, i.e. no po changed its position within the chain. First trials in the automotive industry proved to be far away from that scenario.

To get more detailed information, the production system illustrated in fig.1 is segmented into smaller elements. A segmentation of the whole process into homogenous, subsequent actions of crafting would be ideal. This enforces investments into scanning solutions, po-identification and other information technology. Fig. 1 illustrates the segmentation of production systems at the example of the body shop. Scanning point 1 (SP1) scans initially entering the production segment of a po. The system then stores the po-ID and entering time of all inbound po’s. The entering time of each po describes the real entering sequence. At the end of the production segment the po’s are scanned again (SP2) and the outbound sequence is verified against the inbound one. Detailed localisation of turbulence drivers can be obtained by installing very short production segments.

4.2 TURBULENCE DRIVER EVALUATION

Results of sequence scanning have to be evaluated and documented on a periodical basis. The evaluation must include the analysis of sequence stability over the whole process as well as over its segments. The database structure allows queries regarding:

- number of turbulences over time,
- number of turbulences in production segments,
- number of turbulences under special conditions (damages),
- number of repetitive turbulences.

Detected troublesome segments need to be prioritised. Therefore, functional (failure method effect analysis) FMEA is applied to the evaluation following the procedure:

1. installation of an applicable IT-infrastructure,
2. measurement of real po-sequences,
3. evaluation of sequence data (regarding turbulence frequency in production segments, turbulence repetition, external circumstances),
4. integration of evaluation results into FMEA system,
5. execution of functional FMEA,
6. derivation of countermeasures

One criteria for prioritising is the relation of the total number of processed po’s to the number of turbulences. Weak relations thereby indicate instable processes. After documenting troublesome segments and their functional description, po-contents of turbulences need to be attached to each segment. The statistical frequency of homogenous contents has to be considered. The frequency of turbulences and homogenous contents indicates instabilities that possibly lead to single resources. Another investigation must regard po-contents of turbulences and
their predecessor-successor-combination in terms of critical combinations. Such frequencies have to be assigned to the segments as well. A ranking of critical combination has to be highlighted and checked for repetition.

An additional FMEA criteria is rework. Po’s of rework have to marked and documented in the database as well. The frequency of rework in certain segments has to be evaluated and the results have to be integrated into the FMEA system. Furthermore technical details of rework are to be analysed. The segment of rework origin and execution do not need to comply. Thus, the origins of rework have to be located and considered.

In executing the FMEA algorithm, connections between turbulence reasons, caused by technical or organisational aspects can be found. Organisational uncertainty through the combination of critical po-contents can be reduced. This is subject to detection of critical po-combinations and considering them as new elements of the sequencing criteria area $H_1$. Triggering such contents can then be avoided in future sequencing, while sequence analysis is still in progress. Detected instability of the production resources can be applied to technical redesign of the relevant machines.

5. CONCLUSIONS

At this point, a setup of turbulence counter measures is not yet possible. Our approach presents an organisational and functional solution to locate turbulence drivers. That is why we present an overall algorithm to initially analyse a production system including FMEA Standards. The approach is to be perceived as a pilot trial to allocate the main goals of the pearl chain ppcs of productive and organisational resources. Productive resources, such as single machines, transportation systems or production modules can locally be observed for process stability. Organisational measures of sequencing can be optimized through the limitation of sequencing possibilities. Flexibility and product variance as elements of business strategy can still be sustained. Enhancing these factors in terms of competitiveness, and adapting the production systems is subject of future research.
REFERENCES


Alternative strategies for Vendor Managed Inventory systems in single-vendor/multi-buyer supply environments

M. Falivene, R. Iannone, S. Miranda* and S. Riemma
Dept. of Mechanical Engineering
University of Salerno
Fisciano, Salerno, 84084, Italy

ABSTRACT

The Vendor Managed Inventory (VMI) represent one of the most pursued forms of collaboration which can be established in a supply network. In a VMI system the vendor manages the stocks in the warehouse of the customer and chooses when deliver a new stock of products and the dimension of this stock. In such context, the present research work proposes, through a comparative analysis of the deriving advantages, the application of provisioning management policies alternative to the traditional ones, with particular attention to the single-vendor and multi-buyer environments. The comparison, based on the determination of the total costs of the supply process, has been carried out through the use of opportually developed simulation models. The analysis of the obtained results has evidenced profitable aspects, however conditioned by the interactions existing between the actors of the supply network.

1. INTRODUCTION AND LITERARY REVIEW

The international competitive scenario is even more monopolised by multinational companies which compete among them through a close network of virtual (and often temporary) structures, specific for each typology of product: the so called “supply networks”. The effectiveness of such structures is fundamental for the control and the reduction of the total costs but requires a strong undertaking of all the participants in order to make concrete, also in operative terms, the collaboration described only at a strategic level. Among the different forms of collaboration which can be established, the sharing of data concerning the demand forecasts is one of the most interesting for companies and researchers in the field of operation management. During the years the simple sharing of information has become even closer, up to arrive to Vendor Managed Inventory (VMI) systems.

The choice of a VMI type relationship requires, often, a preliminary analysis targeted to demonstrate its convenience in comparison with traditional stock management policies. The variables which configure a comparison scenario, however, are so numerous and articulated that the evaluation is a problem of significant complexity. For this reason the study of policies which provide for a strong integration between buyer and vendor has been largely discussed in literature.

Many studies have concerned a comparison between the VMI and the Consignment Inventory (CI) policy. Gumus et al. [1] analyse CI in deterministic conditions within a SC composed by one Vendor and one Buyer, deepening when the CI is suitable for the Vendor, for the Buyer or for both of them. The obtained results can be extended also for the VMI when the delivery cost becomes significant. Other researches have compared VMI and CI systems with traditional policies and proposed analytical or heuristic methods to solve logistic problems of the SC. Bichescu et al. [2] studied the effects of the moving of the power to decide in a VMI relationship between Vendor and Buyer. They have analysed, in a mono-product single-vendor/single-buyer context, three scenarios: Powerful Vendor, Powerful Buyer and Equally Power. The results show that the best performances are reached in the Powerful Vendor scenario. Raa and Aghezzaf [3] have studied, in deterministic conditions, the problem of the determination of the distribution plan which minimises holding and delivery costs, in a mono-product single-vendor/multi-buyer context, proposing a heuristic method which furnishes reasonable solutions. Al-Ameri et al. [4]
have analysed and compared several algorithms for the solution of the production, distribution and stock management whole problem, with different approaches.

Southard and Swenseth [5], through a specific simulation tool, have compared different distribution policies in a VMI mono-product multi-buyer context in the agricultural sector, to demonstrate the benefits deriving from the use of a Technology Enabled VMI. The products are stored directly on the means of transport (milk, fuel, etc.) in such a way as to manage the deliveries dynamically. The results, evaluated in 4 different scenarios, have demonstrated that VMI is better than the traditional management and that the costs for the implementation of the technology are largely covered by the obtainable benefits. This also considering the wide spreading of Internet, which has allowed to further reduce the cost of information transferring and sharing within a supply chain, as analysed by Disney et al. [6].

Bertazzi et al. [7] have studied the overall problem of materials production, distribution and warehousing in a VMI context. The study has been carried out with two management policies, called Order-up-to level policy and Fill-fill-dump policy. The solution of the problem has been obtained through its decomposition in distribution and production sub-problems. The results show that in general the VMI produces less costs than the traditional BMI management. This last paper shows several analogies with our study, nevertheless important differences, concerning the proposed management policies and the solution technique, subsist.

The present paper proposes two alternative management techniques, tested in a virtual environment for a mono-product Single-Vendor/Multi-Buyer (SV/MB) scenario and compared with the traditional Buyer managed Inventory (BMI) scenario, in deterministic and stochastic conditions, by varying the production saturation of the Vendor and the levels of stocks of the Buyers. The solution is searched with a real-time approach: exploiting the monitoring data coming from an EDI system it calculates dynamically order release dates and quantities to be produced and delivered, in order to minimise the stocks of the Vendor. Moreover, the cost function includes also the obsolescence cost and the backlog cost, in addition to the other logistic costs already considered in the model of Bertazzi et al. Table 1 reports a comparison between the most recent works on the matter.

### Table 1: Comparison of recent papers on VMI

<table>
<thead>
<tr>
<th>Paper</th>
<th>Vendor</th>
<th>Buyer</th>
<th>Demand</th>
<th>Production Problem</th>
<th>Transport Capacity</th>
<th>Buyers Warehouse</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bichescu (2009)</td>
<td>Single</td>
<td>Single</td>
<td>Stochastic</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>Impact of channel power</td>
</tr>
<tr>
<td>Raa (2009)</td>
<td>Single</td>
<td>Multi</td>
<td>Deterministic</td>
<td>No</td>
<td>Limited</td>
<td>Limited</td>
<td>Inventory routing problem</td>
</tr>
<tr>
<td>Our paper</td>
<td>Single</td>
<td>Multi</td>
<td>Stochastic</td>
<td>Yes</td>
<td>Unlimited</td>
<td>Limited</td>
<td>Analysis of alternative policies of management compared to traditional EPQ/ROP. Integrated solution.</td>
</tr>
</tbody>
</table>

This paper offers as following a formalisation of the problem and the objective function (section 2) and illustrates, basing on them, the proposed policies (section 3). The experimentation (section 4) shows interesting results and allows, in the conclusions (section 5), to define application fields of these policies and hints for further studies.

### 2. SCHEMATIZATION AND PROBLEM FORMULATION

The present work aims to investigate the logistic problem of the VMI relative to a SV/MB scheme, in which the Vendor manages the stocks of its product(s) directly in the warehouses of the Buyers. It has to guarantee a fixed level of service (previously negotiated with the Buyers) in presence of specific volume constraints in the Buyers warehouses. The scenario of the study can be schematised as in Figure 1.

The parameters characterising the Vendor are:

- \( J \): typologies of realised products \( (j \in [1, \ldots, J]) \);
- \( p_j \): production rate for the \( j \)-th product [units/time];
- \( T_{Sj} \): setup time necessary to start the production of the \( j \)-th product [time];
Alternative strategies for Vendor Managed Inventory systems in single-vendor/multi-buyer supply environments

\[ Q_{\text{max}} \] maximum production lot [units];  
\[ Sc_j(t) \] current stock of the j-th product in the vendor’s warehouse [units]; 
\[ ROP_j \] production reorder point of the j-th product in the vendor’s warehouse [units].

The parameters characterising the Buyers are:

- \( I \) number of buyers (\( i \in [1,...,I] \));
- \( ROP_{ij} \) reorder point of the j-th product at the i-th buyer [units];
- \((S,s)_{ij}\) max and min level of stocks of the j-th product at the i-th buyer [units];
- \( Sc_{ij}(t) \) current stock of [units];
- \( d_{ij} \) demand rate for the j-th product at the i-th buyer [units/time];
- \( \sigma_{ij} \) standard deviation of the demand for the j-th product of the i-th buyer [units/time];
- \( q_{ij} \) quantity of j-th product delivered to the i-th buyer [units];
- \( e_{ij} \) quantity of j-th product expired to the i-th buyer [units];
- \( b_{ij} \) quantity of backlog of j-th product at the i-th buyer [units].

The parameters characterising the product are:

- \( C_{Sj} \) setup cost per time unit of the j-th product [€/time];
- \( C_{Hj} \) holding cost of the j-th product [€/unit ·time];
- \( C_{Bj} \) backlog cost of the j-th product [€/unit];
- \( C_{Ej} \) expiring cost of the j-th product [€/unit].

The definition of the parameters concerning each actor characterises the scenario of the study. The modality of production and provisioning management affects significantly the performances of the whole supply system. In fact, given a production policy \( \Psi_h \) and a provisioning policy \( \Phi_k \), the total supply costs will be calculated through the following formula:

\[
CT(\Psi, \Phi) = CTs(\Psi) + CTh(\Psi, \Phi) + CTe(\Psi, \Phi) + CTb(\Psi, \Phi) + CTt(\Phi) 
\]

where:

\[
CTs = \sum_j T_j \cdot C_{Sj} \quad \text{Total setup cost} 
\]

\[
CTh = \sum_j \frac{Ch_j}{T} \int_0^T Sc_j(t) dt + \sum_{i,j} \frac{Ch_j}{T} \int_0^T Sc_{ij}(t) dt \quad \text{Total holding cost} 
\]

\[
CTe = \sum_{i,j} e_{ij} \cdot C_{Ej} \quad \text{Obsolescence cost} 
\]

\[
CTb = \sum_{i,j} b_{ij} \cdot C_{Bj} \quad \text{Shortage cost} 
\]

\[
CTt = C_{js} + D \cdot C_{sw} \quad \text{Transport cost} 
\]
The transport costs are constituted by a fixed part \( (C_{\text{fix}}) \), due to the use of transport means, and a variable part \( (C_{\text{var}}) \), which depends on the covered distance \( D \) and, then, on the chosen path.

Aim of the paper is to establish and compare the advantages, in terms of total supply costs (CT), deriving from the application of a new management policy \( (\Psi_h, \Phi_k) \) in a SV/MB with VMI scenario. The three management policies tested through simulation techniques are: (1) Traditional EPQ (Economic Production Quantity); (2) Lot Merging; (3) Multi-Client EPQ. They will be illustrated in the following section.

3. THE MANAGEMENT POLICIES

3.1. TRADITIONAL EPQ (EPQ)

This model calculates the production lot through the EPQ method. In this case the vendor produces the lot which minimises its unit management costs. Subsequently, every time that a buyer reaches its reorder point \( (ROP_{ij}) \), the quantity necessary to restore the stocks in its warehouse \( (q_{ij}) \) is dispatched. The vendor starts a new production lot when the level of its internal stock reaches the \( ROP_j \) value.

The stock absorption into the vendor warehouse presents a step behaviour (Fig. 2 a): the vendor send the lot to the buyer when it reaches its reorder point \( ROP_{ij} \) (A). Of course the lot, to arrive to destination, will take a time (B) depending on the distance between the two actors. The delivered quantities \( q_{ij} \) are calculated in such a way as to restore the buyer’s stocks to their maximum limit \( S_{ij} \) (C). When the vendor reaches its reorder point \( ROP_j \) (D), it starts the production of a new lot.

The reorder point of the vendor is obtainable from the following formula:

\[
ROP_j = T_{rj} \cdot \sum_{i} d_{ij}
\]

where \( \sum_{i} d_{ij} \) is the demand rate covering all the buyers for the \( j \)-th product and \( T_{rj} \) (reorder time) is equal to:

\[
T_{rj} = T_{s_j} + T_{p_j}
\]

In (8) \( T_{p_j} \) is the time for the production of the lot. The reorder point of the buyers can be calculated as following:

\[
ROP_{ij} = d_{ij} \cdot T_{\text{trasp}}
\]

where \( T_{\text{trasp}} \) is the transportation time.

3.2. LOT MERGING (LM)

This model determines, through an optimisation of the unit costs of the vendor, which lots it is convenient to produce together with the lot of the buyer which has reached its reorder point.

For a better understanding of the LM policy, the following example can be helpful: Buyer 1 has reached its reorder point and so the vendor has to produce the quantity \( q_{1i} \) necessary to restore the level of stock of that buyer. The model evaluates the convenience of producing, together with \( q_{1i} \), the quantities necessary to restore the stocks of the other buyers, by estimating the total unit cost sustained by the vendor. The scheme of Fig. 2b shows the stock tendency of the vendor and two buyers for the LM model. In this case the model has evidenced the convenience of producing also for Buyer 2.

The ROP of the \( i \)-th buyer is calculated as:

\[
ROP_{ij} = (T_{s_j} + T_{p_j} + T_{\text{trasp}}) \cdot d_{ij}
\]

where \( T_{p_j} \) is the time spent to produce the total quantity \( Q_j \) for the buyers belonging to the \( \beta \) set and is given by:

\[
T_{p_j} = \frac{Q_j}{p_j} = \frac{\sum_{i=1}^{\beta} q_{ij}}{p_j}
\]
Alternative strategies for Vendor Managed Inventory systems in single-vendor/multi-buyer supply environments

Figure 2: Stock dynamics

(a) EPQ (b) LM Policy (c) MC Policy

The quantity to be produced for the i-th buyer is given by:

\[ q_{ij} = (S_{ij} - S_{ij}(t)) + (T_{s_j} + T_{p_j} + T_{transp}) \cdot d_{ij} \]  

(12)

To determine the set b of the merged lots, an iterative algorithm, which evaluates the unit cost of the product in function of the level of merging, has been developed. The calculated cost includes set-up, holding during the production and in the buyers warehouses, backlog, transportation. The formula of the unit cost \( Cu \) is the following:

\[ Cu(Q_j) = \frac{C_{fix}}{Q_j} + C_{h_j} \cdot \frac{Q_j}{2} + \frac{C_{set}}{Q_j} + C_{var} \cdot \frac{D}{Q_j} + C_{ch_j} \]  

(13)

Fig.3 shows the flow diagram of the LM model algorithm. It is constituted by the following steps: (step 1) Identifying of the lot \( Q_j \) of the x-th buyer which has reached first its reorder point; (step 2) Determination of the total unit cost sustained by the vendor if it produces the lot \( Q = Q_j \); (step 3) Determination of the total unit cost if \( Q \) is increased with the quantity \( Q_i \) of the buyer in the nearest position; (step 4) If the unit cost calculated with \( Q + Q_i \) is lower than the cost calculated with \( Q \) the algorithm goes to step 5, otherwise it goes to step 6; (step 5) The new production lot is: \( Q = Q + Q_i \); (step 6) The set of buyers in the outskirt of the x-th buyer is reduced by the i-th element; (step 7) If the set of buyers near to the x-th is empty, the algorithm goes to step 8, otherwise it returns to step 2; (step 8) The final lot \( Q \), given by the sup of the merged lots, is furnished as output of the algorithm.

3.3. EPQ MULTI-CLIENT (MC)

The MC model determines the production lot basing on the mean total demand of the buyers. The production starts when the sum of the quantities required by the buyers become equal to the EPQ calculated considering setup, production and transportation times. The time necessary to produce the lot is given by this formula:
The single lots can be calculated as following:

\[ q_{ij} = (S_u - S_{c_i}(t)) + (T_s + T_p + T_{asw}) \cdot D_j \]  \hspace{1cm} (16)

In Fig. 2c is represented the stock dynamic of vendor and two buyers with the MC policy. Once the lots have been produced, they are delivered to the respective buyers in order to restore their level of stock to the \( S_{ij} \) value.

4. EXPERIMENTATION

The convenience of the proposed policies, in comparison with the traditional logic \((\Psi_0, \Phi_0)\), has been evaluated through the performance index \( R \), obtained with the following relation:

\[ R = \frac{C_T(\Psi', \Phi', P_0)}{C_T(\Psi_0, \Phi_0)} \]  \hspace{1cm} (17)

If \( R \) is negative, the tested policy is economically more convenient than the EPQ. The experimentation, carried out by means of opportunely developed simulators, consists of the following points:

1. comparison between the scenarios in deterministic conditions by varying the saturation of the vendor and the mean level of the buyers’ stock;
2. Analysis of the effects of point 1 in aleatory conditions;
3. Analysis of the effects of safety stocks in aleatory conditions.

The comparison has been carried out by varying the following parameters:

1. \( p_j \cdot \sum d_i \) \hspace{1cm} \text{ratio between production rate and total demand rate;}
2. \( \sum S_i / Q_i \) \hspace{1cm} \text{ratio between maximum level of stock and EPQ of the vendor.}

4.1. DETERMINISTIC SCENARIO

The tested SV/MB scenario has been configured for the production of an unique product \((J=1)\), in any quantity \((Q_{max} \to \infty)\), to be delivered to 5 buyers \((I=5)\). The values of the parameters used for the test are the following:

<table>
<thead>
<tr>
<th>Parameters</th>
<th>( p_j / \sum d_i )</th>
<th>( \sum S_i / Q_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned values</td>
<td>1.9 2.2 2.5 1.1 1.6 2.2 2.6</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4a shows that the convenience of the LM model grows with high values of \( S/Q \) and \( p/d \) ratios. This effect is principally due to the increase of the costs of the traditional model, while they remain steady in the proposed model. Fig. 4b evidences that with the MC model the \( R \) index has an opposite behaviour; in fact, the convenience grows when \( S/Q \) decreases. The \( p/d \) ratio, on the contrary, has not a significant effect. This first test evidences that the two policies have complementary convenience domains: therefore, better results than the EPQ can be obtained in the entire experimentation domain.
Alternative strategies for Vendor Managed Inventory systems in single-vendor/multi-buyer supply environments

4.2. ALEATORY SCENARIO

In this scenario the standard deviation ($\sigma$) of the buyers demand has been introduced. The simulation has been conducted with the two best configurations of the previous scenario. This test aimed to verify if the results obtained in deterministic conditions remain good also with aleatory demand. For each configuration the values of 20% and 40% of the $\sigma/d$ ratio have been considered.

Table 2 shows mean and standard deviation of the $R$ index in function of the $\sigma/d$ ratio in the best configuration of the MC model: the policy remains convenient (high value of $R$) and also steady (reduced standard deviation). With the same parameters, also the LM policy has been tested, in order to verify if the results improve in presence of aleatory conditions. The results obtained in this configuration with the LM model confirm that this system, even if rather convenient, is unstable because the standard deviation presents very high values.

Vice versa, when the parameters are the best for the LM policy, the results are opposite. In fact, Table 3 shows clearly that the MC model is not convenient and unstable, while the LM model improves its results also reducing their dispersion.

4.3. ALEATORY SCENARIO WITH SAFETY STOCKS

A further test has been conducted to consider the effects of the safety stock ($ss$) on the EPQ configuration with the lowest level of service ($p/d=1.9$ e $S/Q^*=1.1$). In this case $\sigma/d$ has been fixed to 40%, while the minimum stock level of the buyers varies from 0 to 10% of the maximum level. The obtained results are shown in Fig.5.

Also in this case the MC system is convenient and stable (Fig. 5a). The presence of the $ss$ not affects significantly the results; this is due to the mean level of stock that, with this policy, is always high because they are...
supplied before have reached their reorder points. The LM policy, with demand variability and safety stocks, produces very negative results (Fig. 5b). This is due to the fact that the safety stock anticipates the reorder points of the buyers and so the required quantities are reduced. The system is therefore forced to make small mergings which do not allow the subdivision of transport and production fixed costs on a significant number of pieces.

5. CONCLUSIONS

The analysis of the data collected during the simulation has evidenced many advantages of the proposed policies. Between the alternatives, the MC policy is surely the most convenient and stable. For this reason, given the high variability of the production systems, it is potentially the most suitable for a large use. However, this policy is not ever appropriate: in fact, when production rate and mean stock are high the risk of obsolescence grows considerably and the LM policy become preferable. The analysis has also evidenced that, when the product value is low and the level of service is required high, the best policy is the traditional EPQ which, however with higher total costs, guarantees a level of service the other policies are not able to guarantee.

From the observation of the obtained results, the $S/Q$ ratio results the factor which mainly affects the convenience of one or another policy. In fact, when such ratio varies, the total costs, and therefore the $R$ index, change remarkably. On the contrary, the $p/d$ ratio is not significant, with the exception of the LM model in which an increase of the produced quantity determines an notable variation of the backlog and the obsolescence costs.

As all the simulation models, also these suffer from the effects deriving from the interactions with the other elements of the logistic chain which the system belongs to. Nevertheless, with the results obtained by the different simulations carried out in both static and dynamic conditions, the proposed policies represent a valid support to the management of the VMI systems.

REFERENCES


Development of an Online Supplier Selection System for a Turbine
Manufacturing Company

1L.Siva Rama Krishna*, 2G.Ranga Janardhan, 3C.S.P Rao

1 Dept. of Mechanical Engineering, University College of Engineering(A), Osmania University,
Hyderabad, Andhra Pradesh, India.

2 Dept. of Mechanical Engineering, JNTU College of Engineering,
Kakinada, Andhra Pradesh, India,

3 Dept. of Mechanical Engineering, National Institute of Technology,
Warangal, Andhra Pradesh, India.

ABSTRACT

In today’s highly competitive global market scenario, manufacturing companies are constantly under pressure to identify ways of optimizing the cost of production. Material purchasing cost plays an important role in deciding the manufacturing cost. The material purchasing cost depends upon the supplier selection process which is a complex process involving qualitative and quantitative criteria. Thus an intelligent supplier selection system is required by managers which help them in identifying the right combination of suppliers so that the total cost of procurement is minimized. The present work discusses the development of an online supplier selection module for a turbine manufacturing company located in India. The objective of the material purchasing department of the user industry is to minimize the total cost of procurement of material by determining the optimum quantity of material to be procured from each supplier satisfying the demand. The system is developed in two phases. The first phase involves the development of a supplier selection system. The system helps in identifying the right combination of suppliers from the available suppliers for a given material. While selecting the suppliers along with the cost of material to be procured from each supplier the minimum and maximum capacities of the suppliers are also taken into consideration. Scenarios are generated with supplier combinations in the form of sets. A heuristic search technique is developed for identifying the best set of suppliers. The second phase involves integrating it with World Wide Web (WWW) which improves the flexibility of the system so that it can help in procuring the material from suppliers distributed around the globe. The primary motivation for adopting this technology is to enable the manufacturers to establish effective supply chains in less time and less cost as compared to traditional procurement methods.

1. INTRODUCTION

In today’s highly dynamic environment, companies are facing increasingly intense competition to reduce their cost of production. Material purchasing cost plays a vital role in determining the total cost of production. The material purchasing cost depends upon the supplier selection process, which is a decision making process to identify and select the best suppliers using certain set of criteria [1]. The traditional supplier selection processes involves all the activities from needs definition to contract management [2]. These processes are usually executed on stand-alone software systems, which are normally used in personal computers or local networks. Due to globalized market conditions, the suppliers may be located at geographical different locations thus making it necessary to use the potential of internet. It is acting has a global link between company’s customers, managers and suppliers [3] The internet is the technology that has been used for all business-to-business relationships [4]. It has led to new opportunities to select suppliers online, due to wider search base, broad and fast information exchange and lower transaction cost [5]. The wide use of the Internet has enabled buyers to locate large number of suppliers and has

* Corresponding author: Tel.: (091)-040-64572017 ; E-mail: s_r_kris@rediffmail.com
facilitated suppliers to let buyers know of their existence. The Internet has become an e-marketplace where buyers and sellers interact electronically [6].

The present paper is an extension of the one we presented in our earlier work [7], where we discussed about the development of a web integrated decision support system for machine scheduling and inventory management. In particular, in this work we discuss the development of online supplier selection system for a turbine manufacturing industry. The system is developed in two phases, the first phase is the development of a intelligent supplier selection system as per the requirements of user industry and the second phase involves in making the system web enabled.

Rest of this paper is organized as follows: Section 2 presents a detailed description of the problem. Section 3 discusses the development of intelligent supplier selection system. This includes assumptions and the heuristic developed for solving a real-time large scale supplier selection problem with a numerical illustration. The web enabling of the supplier selection system is discussed in section 4. Finally section 5, summarizes the work presented.

2. PROBLEM DESCRIPTION

The problem is described from the perspective of a typical steam turbine manufacturing industry located in India. The company manufactures different types of turbines of varying capacities in a year depending on the customer orders. Each customer order is identified by a unique number known as work order number (WONO). Each work order consists of several subassemblies identified by a unique number termed as product group main assembly (PGMA) and each product group consists of individual parts identified by part number (PARTNO). Each turbine comprises of 50 PGMAs and each PGMA has 50 parts on an average and each part undergoes 25 to 125 operations. For example, the outer casing, the exhaust hood and the valve chest are important sub-assemblies for any steam turbine. The number of operations for outer casing is around 125 operations, exhaust hood has around 45 operations, and valve chest has got 25 operations. The typical product structure of a steam turbine [7] is shown is illustrated in Figure 1.

![Figure 1: Product Structure of Steam Turbine [7]](image)

The key task of the material purchasing department of the user industry is to minimize the total cost of procurement of material by determining the optimum quantity of material to be procured from each supplier satisfying the demand, the minimum and maximum capacities of the supplier. In this paper we have concentrated on the cost of procurement of raw material only for the parts which are manufactured in the plant. The raw material may be common for different parts belonging to the same PGMA and work order or different PGMA and work order. To help the material purchase department in achieving their task an online supplier selection system is developed.
The system is developed in two phases. The first phase involves the development of supplier selection module. A heuristic is developed for identifying the best combination of suppliers so that the total cost of procurement of raw material is minimized. The second phase involves integrating it with World Wide Web (WWW) which improves the flexibility of the system so that it can help in procuring the material from suppliers distributed around the globe.

3. SUPPLIER SELECTION MODULE – PHASE-I

In this section we discuss the heuristic adapted in solving the supplier selection problem of the user industry under investigation. The supplier selection is done taking into various factors like minimum and maximum capacities of material and partial supply of raw material of a supplier. The following assumptions are made for the heuristic procedure adopted.

3.1. ASSUMPTIONS

1. Quantity discounts are not considered
2. The entire replenishments of each period must be available at the beginning of periods.
3. It is assumed that the supplier supplies the material only in steps of equal increments from minimum to maximum capacities. For e.g., if the minimum is 20 and maximum is 60 and increment is 5, then he supplies material in batches of 20, 25, 30, …40, but not in batches of 21, 22 etc. This is an important assumption, because it takes care of partial supply of material by a supplier.
4. The demand in one period is shared between suppliers, so that the best combination of suppliers can be selected so as to minimize the cost of procurement.
5. The capacities of suppliers may be different.
6. The only cost considered is the cost of raw material from each supplier, which is different for each supplier.

3.2. PROPOSED HEURISTIC FOR SUPPLIER SELECTION

Step 1: The demand and the supplier details like name of the supplier, minimum and maximum capacities, cost of procurement of material from the supplier and steps of increments in which he can supply are taken as input for generating the supplier sets.

Step 2: Based on the increment, the supplier sets are generated between the minimum and maximum capacities for each supplier.

Step 3: The combination of supplier sets are generated in random satisfying the demand and minimum and maximum capacities of the supplier. The cost of procurement of each set is calculated.

Step 4: In this step the best set is identified by applying the least cost algorithm.

Least cost algorithm:

i. Accept the first generated random set and the cost of the set.
ii. From the second set onwards, accept the set if the cost of the present set is less than the previous set, otherwise reject it.
iii. Continue until the minimum cost set is identified. The best set is one, after which the cost of remaining sets is more.

Step 5: Based on the identified least cost set, the system selects the suppliers and also the quantity of material that has to be procured from each supplier. Then an order is placed to the corresponding supplier based on the lead time of the supplier.

3.2.1 NUMERICAL ILLUSTRATION

The functioning of the above algorithm is illustrated with an example in this section.

Step 1: Demand: 42 tons Increment: 2
The supplier details which act as an input for the algorithm is shown in Table 1.

### Table 1: Supplier details

<table>
<thead>
<tr>
<th>Supplier Name</th>
<th>Minimum Quantity in tons</th>
<th>Maximum Quantity in tons</th>
<th>Cost of raw material in INR per Ton</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>40</td>
<td>5000</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>40</td>
<td>6000</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>40</td>
<td>7000</td>
</tr>
</tbody>
</table>

**Step 2:** Possible supplier sets generated between minimum & maximum quantity with an increment of 2 are shown in Table 2.

### Table 2: Supplier sets

<table>
<thead>
<tr>
<th>Supplier Name</th>
<th>Supplier Sets generated between min. and max. capacities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 0</td>
</tr>
<tr>
<td>B</td>
<td>10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 0</td>
</tr>
<tr>
<td>C</td>
<td>10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 0</td>
</tr>
</tbody>
</table>

Note: zero value in the above set indicate material is not procured from the supplier.

**Step 3:** For the above sets, 64 random sets are generated satisfying the demand and minimum and maximum capacities of the supplier. A Sample of generated supplier sets are shown in Table 3 below.

### Table 3: Random supplier sets generated

<table>
<thead>
<tr>
<th>SET NUMBER</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>TOTALCOST in INR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>14</td>
<td>18</td>
<td>260000</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>16</td>
<td>16</td>
<td>258000</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>18</td>
<td>14</td>
<td>256000</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>20</td>
<td>12</td>
<td>254000</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>22</td>
<td>10</td>
<td>252000</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>10</td>
<td>22</td>
<td>264000</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>12</td>
<td>20</td>
<td>262000</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>10</td>
<td>20</td>
<td>260000</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>12</td>
<td>18</td>
<td>258000</td>
</tr>
</tbody>
</table>

For random set generation, a graph is plotted between the supplier set No. and the total cost indicating the trend of total cost for different combinations of supplier sets. Figure 2 shows the graph of random set generation.

![Figure 2: Random Set generation](image-url)
Step 4: Sets generated after applying the least cost algorithm. The number of sets reduced to 21 sets. Table 4 shows these sets.

Table 4: Least cost supplier sets generated

<table>
<thead>
<tr>
<th>Set Number</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total Cost in INR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>14</td>
<td>18</td>
<td>260</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>16</td>
<td>16</td>
<td>258</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>18</td>
<td>14</td>
<td>256</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>20</td>
<td>12</td>
<td>254</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>22</td>
<td>10</td>
<td>252</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>20</td>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>18</td>
<td>14</td>
<td>18</td>
<td>10</td>
<td>248</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>16</td>
<td>10</td>
<td>246</td>
</tr>
<tr>
<td>25</td>
<td>18</td>
<td>14</td>
<td>10</td>
<td>244</td>
</tr>
<tr>
<td>27</td>
<td>20</td>
<td>12</td>
<td>10</td>
<td>242</td>
</tr>
<tr>
<td>28</td>
<td>22</td>
<td>10</td>
<td>10</td>
<td>240</td>
</tr>
<tr>
<td>31</td>
<td>14</td>
<td>28</td>
<td>0</td>
<td>238</td>
</tr>
<tr>
<td>32</td>
<td>16</td>
<td>26</td>
<td>0</td>
<td>236</td>
</tr>
<tr>
<td>33</td>
<td>18</td>
<td>24</td>
<td>0</td>
<td>234</td>
</tr>
<tr>
<td>34</td>
<td>20</td>
<td>22</td>
<td>0</td>
<td>232</td>
</tr>
<tr>
<td>35</td>
<td>22</td>
<td>20</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>36</td>
<td>24</td>
<td>18</td>
<td>0</td>
<td>228</td>
</tr>
<tr>
<td>37</td>
<td>26</td>
<td>16</td>
<td>0</td>
<td>226</td>
</tr>
<tr>
<td>38</td>
<td>28</td>
<td>14</td>
<td>0</td>
<td>224</td>
</tr>
<tr>
<td>39</td>
<td>30</td>
<td>12</td>
<td>0</td>
<td>222</td>
</tr>
<tr>
<td>40</td>
<td>32</td>
<td>10</td>
<td>0</td>
<td>220</td>
</tr>
<tr>
<td>41</td>
<td>10</td>
<td>0</td>
<td>32</td>
<td>274</td>
</tr>
<tr>
<td>53</td>
<td>10</td>
<td>0</td>
<td>32</td>
<td>284</td>
</tr>
</tbody>
</table>

Note set numbers 41 and 53 are shown in the above table to indicate that after set 40 other sets have higher cost. Figure 3 shows the graph representing the trend after applying the least cost algorithm.

![Figure 3: Least Cost algorithm](image)

Step 5: From the above table set number 40 is identified as the least cost set and suppliers A and B are selected. Supplier A supplies 32 units and supplier B supplies 10 units satisfying the demand and the least cost is Rs. 22000.
4. WEB ENABLING OF SUPPLIER SELECTION SYSTEM

The main advantage of web enabling any application is its inheritance of client server architecture for distributed decision making. The present supplier selection system is web enabled using Java 2 Enterprise Edition (J2EE) techniques. The web pages are developed using Java Server Pages (JSP). The JSP pages contain all the required graphical user interfaces (GUI’s) between the client and the server. It allows authorized users to select the desired parameters for supplier selection viz., entering the details of the supplier like type of material, minimum and maximum capacities, cost of raw materials and other relevant details. The Security management is achieved through password authentication and firewalls. It provides two types of authorization levels, namely: manager and Supplier. Different users have different login and passwords. The developed system adopts three-layer architecture.

4.1 THE STRUCTURE –IMPLEMENTING THE THREE LAYER ARCHITECTURE

The system is developed based on the client /server approach with the three layer architecture. The three layers are presentation layer, application layer and database layer. Figure 4 shows the structure of the method, which enables the user to execute the system from any remote location.

The first layer is the presentation layer which consists of web browsers. The first layer allows the authorized users to interact with the system located at a remote place.

The middle layer is the application layer. It consists of two logical layers, the web application program and the standalone application program. The web application program consists of web service clients and struts framework. The web service client communicates with the web services present in the standalone application program. While the struts framework is a standard for developing well-architected web applications. It will process requests from the client and interacts with the web service client. The standalone application program consists of the web service and the Visual Basic (VB) program. The Web services execute the VB program and sends the response to the web service client. The VB program consists of code for supplier selection heuristic.

The third layer is the database layer. The database layer consists of relevant data pertaining to the supplier details like name of the supplier, minimum and maximum capacities of supplier etc. Oracle 10G is used as Database Management System.

![Figure 4: Three tier architecture of online supplier selection system](image)

4.2 USER INTERFACES OF ONLINE SUPPLIER SELECTION SYSTEM

The online supplier selection system has user friendly graphic interfaces which help various clients in operating the system with ease. Some of them are discussed below.
**Home Page:** The home page is designed for two types of users namely the manager and supplier. Suppliers are allowed to register and enter their details. Managers are allowed to access all the web pages and also to modify the data. Figure 5 represents the home page.

![Home page of Online supplier selection system](image)

**Figure 5:** Home page of Online supplier selection system

**Supplier Details Page:** After the supplier logins, in this page he has to enter various details like type of material he supplies, minimum and maximum capacities, cost of procuring material and increment of material supply. Figure 6 illustrates the supplier details page.

![Supplier details page](image)

**Figure 6:** Supplier details page

**Supplier Selection Page:** This page displays the best suppliers set after executing the supplier selection heuristic discussed in section 3.2. The access to this page is given to the manager of the material purchasing department. Figure 7 illustrates the supplier selection page.
5. SUMMARY

In this paper we have discussed the development of an online supplier selection system for a turbine manufacturing industry. The system helps the decision makers in identifying the right combination of suppliers satisfying the demand in a period at a least cost. The developed system takes care of the minimum and maximum capacities of the suppliers while determining the best combination of supplier sets.

A heuristic is presented to identify the best supplier set such that the total cost of procurement is minimum. The developed system also takes care of the partial supply of the material by a supplier. The system is flexible enough to incorporate any changes in the demand, supplier details and number of suppliers. The system is web enabled thus allowing the suppliers located at geographical dispersed places to interact with it. The work presented in this paper helps the managers in arriving at a right decision while selecting the supplier. As supplier selection process is a vide area additional features like ordering cost, logistics etc can be added to increase its performance i.e. improve the capability of the system in meeting real time constraints.

REFERENCES

Handling of Time, Complexity and Dynamics within Supply Chains

Thomas Burghardt¹, Hendrik Jähn, Sascha Häckel,
Department of Economics
Chemnitz University of Technology
D-09107 Chemnitz, Germany

ABSTRACT

Since its development in the Japanese automotive industry, Lean Management has essentially supported enterprises with the elimination of waste. Waste, among others, occurs by means of overproduction, idle times of people and machines (waiting for the next production step), unnecessary transports, an oversized stock of inventory and defective products. Each form of waste represents a consumption of resources without getting any service in return, as the customer is not willing to pay a higher amount for it. Taiichi Ohno (one of the two fathers of the Toyota Production System) already postulated this in his homonymous book: Each form of wastage has to be avoided as possible. The factor time and the control of complexity and dynamics become more and more important, not only out of the Lean Management’s point of view. The Just-in-time (JIT) approach and the connection of activities and processes to continuous running, demand-driven flows of information and material within supply chains (SC), do reflect this change. The flow is understood as a metaphor to whose ideal - which is a continuous, uninterrupted flow - the processes should be orientated. Current analyses deal with models and methods for a continuous and smooth flow of material in SCs. By means of a resolution of “barrages” (production without taking any inventories), idle times can be eliminated, which thereby supports the “Lean”-approach. The basic edition about science discipline of Cybernetics appeared in 1948 by Norbert Wiener. The title of the book is: „Cybernetics or Control and Communication in the Animal and the Machine”. Nowadays, cybernetics is further distinguished, among others, within the systems theory and control engineering. Fundamental ideas and concepts of cybernetics are also applied in economics, e.g. in a cybernetic oriented modelling and an experimental simulation of economic processes. Cybernetics is understood as a formal science of the structure, relations and the behaviour of dynamic systems. Within this article the analysis of formal cybernetic principles occurs for a stable operation in order-related production networks (OPN).

INTRODUCTION AND MOTIVATION

Supply chains obtain the biggest part of their dynamics, but also many of their potentials for success out of the harmonised and accelerated transfer of information. Information consists of data and form the basis of knowledge. From this point of view, information play a key role in modern economics and therefore not least in the supply chains. Because of this, information (or knowledge) [12] is meanwhile described as a factor of production besides the classical production factors out of an economic point of view, which are work, land and asset [11]. If the economic development, including its increase cycles, is considered over the last centuries, this assumption appears to be conclusive. Here, different economic cycles can be identified over the time flow. SCHUMPETER [13] established the term „Kondratieff-Cycle”, for those long-orientated cyclical waves. He found out that such cycles are first of all caused by fundamental, technical and economical innovations. Out of a production- and organization-related perspective, this leads to a basic reorientation. Development and its world-wide consistent implementation of modern communication technique is here interpreted as an own Kondratieff-Cycle. Here, this concerns the 5th cycle, whose start lies in the first half of the 1990s. A basic trend was here justified in the development of a global economy. The first years after the German reunion at the beginning of the 1990s were defined by a bad economic situation, which was also characterized by a downturn for the registration of new vehicles. The worldwide production of cars decreased drastically during that time. At this time, the overall economic situation was rather bad. It is remarkable that during this time of cyclical downturn, Toyota’s profit indeed decreased, but however, not to the extent of comparable business rivals. Therefore, Toyota produced 4.5 Million vehicles in 1991, with a work force of

¹ Corresponding author: Tel.: +49-371-531-35576; Fax.: +49-371-531-26289; E-mail: thomas.burghardt@wirtschaft.tu-chemnitz.de
70,000 employees, Volkswagen with 260,000 employees, but an output of only 3.5 Million vehicles [15]. The competitive pressure of far eastern competitors was the American and European producers increasingly giving a hard time. First of all, a disappointing cost structure was determined for the European car manufacturers.

Within empirical studies, HEIL [9], who had undertaken an inventory regarding the disturbance situation within normal operations in 1992/1993, also got to this point. On the basis of those results, HEIL describes cause-and-effect-relations by means of those results and deduces measures for the suppression of disturbances in enterprises.

In summary, the introduction concludes that the world-wide distribution and use of the internet did not occur accidentally during that time. Cause-and-effect relations are observable. This technology formed the basis for the globalization of economy and the economic improvement connected to it within the fifth Kondratieff-Cycle.

**EMPIRICAL STUDIES FOR OPERATIONAL DISTURBANCES**

Within this chapter, a summary of the results of analyses obtained by HEIL, regarding the causes and effects of disturbances, will be undertaken. The sample tested by HEIL consists of 25 enterprises out of the following branches: Automotive engineering and –supply, chemical industry, electrical engineering and electronics, machine- and plant construction, iron-, plate- and metal processing as well as the optical industry. Due to a comparatively small sample size and the heterogeneity of the consulted enterprises, there was no demand for a representativeness of the results.

Among others, the target was the development of trend statements regarding the disturbance situation within the internal value added process. At the time of the interrogation (1992/93), the enterprises were under competitive pressure. The economic situation in Germany was defined by economic changes already mentioned within the introduction paragraph. This is why the enterprises were first of all asked, which basic strategies were followed for an output increase in the past (the last 10 years), momentarily and which will be being followed in future (within the next 10 years). Before the start of the fifth Kondratieff-Cycle, a focus in the basic strategies of the enterprises could be observed within the area of process control and production engineering. There have been higher investments in PPS or the improvement of plant availability. Sales collapses and declining shares in the market within the first years after the reunion forced the analyzed enterprises to reconsider their basic strategies. Strategies for a decrease of inventory and lead time reductions were held, but the implementation no longer exclusively occurred through the production planning and controlling system (PPS) of the enterprise. The companies started to invest in new holistic strategies (considering the supply chain). They for example realized that the problem of inventory is also a logistical one, which requires the inclusion of procurement, production- and distribution department. For the analyzed enterprises, HEIL observed further adjustments of the basic strategies as a result of the changed competition situation in the short-, middle- and long term. The procurement at the time of manufacturing for example gained higher importance. Another fourteen enterprises were mainly dealing with the optimization of the flow of material by means of structural changes of the process elements. Employee qualification as well as programs for a continuous improvement either became more important or were started again. Besides the newly introduced basic strategies, HEIL investigates the *causes of disturbances* connected to it and with which the enterprises were confronted with. The evaluation clearly shows that intensified requirements of the market regarding the fast and more flexible distribution at a high quality level have direct consequences for the in-house disturbance situation. The reduction of securities within the JIT-projects dominated for the causes of conditions before growing demands for quality and the changed order performance of the customers from the consulted enterprises’ point of view. Therefore, the JIT-concept shall not be considered as cause for disturbance but only as a frame-giving factor, which enforces the disturbance potential. The challenge for the enterprises lies within the problem to find ways to ensure a disturbance-sturdy production despite the increased disturbance potential. Regarding the activity reasons (equipment, people, material, information, control) it has been investigated by means of which individual causes disturbances will be evoked. *Disturbances in control and logistics* were mentioned most frequently by the respondents. Missing parts or lacking capacities of transporting for example belong to the last mentioned complex of causes. During his study, HEIL eventually also concentrated on the *impact of disturbances* for the enterprises. The basis for an effect-related interference suppression strategy is the recognition of single emphases of impact within the value added chain. The influence of disturbances was checked regarding the strategic factors of success: time, quality and costs. The time dimension was represented by the criteria processing time and delivery time. According to experts, success of enterprises is mainly influenced negatively by extensions of processing times caused by disturbances. Therefore, the aspect time seems to be of special importance for the conception of the company’s fault clearance. Interested readers are advised to research reference [9] for further information concerning the results of the survey performed by HEIL.

The cause-and-effect relations considered by HEIL within his study confirm the assumption that the control over the factors time, complexity and dynamics influences the competitive situation within enterprises. This assumption applies for individual enterprises as well as corporate networks. The article presents measures for a handling of
mentioned challenges. Measurements for an operational interference suppression suggested by HEIL will, among others, be applied exemplary to the disturbance management of order-related production networks. Within this context, formal cybernetic principles are highlighted in particular. The target is a disturbance-sturdy operation of OPN. In the following two main chapters, primarily the current state of research regarding OPNs is presented.

ORDER-RELATED PRODUCTION NETWORKS – A CLASSIFICATION

Small and medium-sized enterprises which cooperate in order-related production networks (OPN) are in the focus of research at Chemnitz University of Technology. Networks are classified according to their type of management and the temporal stability of their inter-organisational relations. Within dynamic networks, legally independent enterprises mainly cooperate without rigid dependencies. According to this typology, OPNs are controlled heterarchically and need to be classified dynamical regarding their temporal stability [1]. Reference [2] illustrates this classification with regard to the network type OPN in table 1. The target of an OPN is to use the advantages of the supply chain management in a short time frame to achieve an individual satisfaction of customer requests.

OPERATOR CONCEPT

For the coordination of such added value processes, which mostly focus on single and small serial manufacturing, the operator concept „Extended Value Chain Management“ (EVCM) has been developed [3]. The EVCM is a network management concept for the operation and coordination of OPNs. It is the target of the EVCM to select those enterprises for a production process which are best suitable for a concrete customer order. Interested readers are advised to consider [2], [3] and [4] for further information concerning the genesis of production networks supported by the operator concept.

Tasks of the EVCM within order-related production nets are the selection of competence cells during the generation of production nets as well as the formation, control and monitoring of the operation of the production net. The primary target is a stable operation of the production net. The concepts discussed in reference [5] and [8] for monitoring and control of OPN contain the 3-level conception for monitoring as well as measures for interference suppression within a preventive and reactive disturbance management. Both concepts will be presented in short subsequently. Caused by the current state of realization of the EVCM, it has to be noted at this point that no IT-supported implementation of the presented approaches for monitoring and disturbance management does exist so far.

MONITORING

Experiences out of the project management finally evaluate communication as a project’s most important factor of success [7]. The required information need to be provided to the different actors to the requested extent, also within a production net. For the collection and editing of process information, a 3-level conception is recommended for monitoring [8]. ZSCHORN [8] pointed out that the lowest level also involves the condition monitoring, which is furthermore described as equipment conditioning monitoring. It specifies the constant recording of the state of complex technical plants and systems. Those systems are applied mainly for monitoring of the machines used and for observation and control of transfer lines. At the medium level, also found within the enterprises, monitoring of production processes mainly occurs on the basis of a WIMS and production data acquisition. This task is organized individually within the enterprises and is restricted to the value-added process. For the participating enterprises of the OPN, there is no difference in the processing of a network order in contrast to other orders concerning the operational monitoring. At the highest level of the production networks, the operational network management is responsible for it. The target of monitoring lies within the control of supply agreements by network participants as the basis for the scheduled procedure of the value-added process within the production network. The provision of data for the monitoring is taken over by a distributed WIMS [5]. The essential task of a WIMS, in the context of EVCM, consists in the support of the participants during the coordination of activities within the production net. During the production process, the tasks data collection (current values) and evaluation of data is continually taken over by the monitoring on the network level and provides for an early operation or fast reaction in case of disturbances by means of monitoring figures (target values). The evaluation of data occurs by means of a target-actual comparison in consideration of the admissible tolerance range. The WIMS recognizes, if there is a
disturbance or not because of the target-actual-comparison. This equates to a typical feedback loop, which consists of a system, a controller, the control variables and the feedback (cp. Fig. 1). The supply chain, which consists of the linked value added units, complies with the system. The date of delivery or the delivery reliability are examples for control variables. The disturbance management functions as the controller and, where applicable, changes the value. The opportunity of the disturbance management to exert influence is e.g. by means of a reconfiguration of the net.

Figure1: Feedback Loop

**DISTURBANCE MANAGEMENT**

The disturbance management within the EVCM provides instruments for interference suppression during the entire life cycle of a production net. Within this article, the focus lies on the disturbances, which can occur during the networking. Disturbances of company processes belong to the basic problems of industrial enterprises. Causes for disturbances as well as effects of disturbances need to be considered during any suppression of interference in enterprises [9]. The disturbance management is subordinated to the operational network management and relates to procedures for a removal of disturbances and therefore the recovery of a normal flow of information, work or material [7].

Appropriate measures are presented in the following, whereas the focus in this article is especially on formal cybernetic principles.

**KEY FIGURE DELIVERY RELIABILITY**

Reference [8] discusses the demand of measurable process parameters or key figures, as the decisional base for the disturbance management. Typical key figures of monitoring for ZSCHORN are the date of delivery as well as quantity, quality and price. In reference [8], it is particularly referred to the meaning of the supporting key figures probability of delivery and capability of production, which are summarized in the key figure delivery reliability. Both values are calculated again [3][5] for each enterprise during the formation of the production network and, during the operation of the production net, collected as key figures for disturbance management in the form of continuous surveys for network participants, which will follow during the value-added process [16]. This offers the opportunity to the disturbance management to determine the deviations between the probability of delivery or production of capability in case of a scheduled procedure and the real current value for the network participants. This in turn provides for an early prediction of disturbances of the value-added process [8]. In this context, ZSCHORN furthermore differs between a preventive and reactive disturbance management. In case of an increase of uncertainty within the production network, a respective decrease of delivery reliability will be the consequence. By means of appropriate measures of the preventive disturbance management, the occurrence of disturbances on the enterprise level or at least on the level of the production net can be avoided or their effects can be decreased, so that the real process lies within the admissible range. The essential advantage of the preventive disturbance management on the basis of key figures of the probability of delivery and capability of production consists in an increase of time, which is still available for reaction. Reference [8] suggests to use the time, if applicable, for a rescheduling or reconfiguration of the residual value-added chain, as soon as it is foreseeable that the participant cannot fulfil his supply in due time. If unpredictable disturbances occur during the operation of the production network, measures of the reactive disturbance management ensure a fast reaction in case of deviations.

This procedure occurs similarly to the preventive disturbance management through an application of methods of the network generation (reconfiguration). An example is given by the search of process alternatives, if disturbances cannot be resolved by single network participants and if time buffers are insufficient. In this case, a partial rescheduling can be undertaken by the disturbance management.
DELIVERY BUFFER

With regard to a stable and therefore disturbances-sturdy operation of OPN, FISCHER recognized the delivery buffer as potential reliability factor and has taken it into account during the setting up of networks. The delivery buffers show the maximum allowable delay of the supplier without an expiration of the purchaser’s offer or delays on the part of the purchaser. Therefore, the delivery buffers, which flow into the optimization as time units on the edges of the offer net, are independent from each other [19].

SIMULATION

According to EVERSHEIM, a successful disturbance management requires a current and precise description of the state of the process, a recording and documentation of occurred disturbances and the analysis or evaluation of disturbances [17]. One opportunity of the disturbance management consists in the use of simulations for the determination of the effectiveness of decisions considering possible occurring disturbances and measures necessary for a prevention regarding the compliance with deadlines [18]. Before a simulation can be undertaken, an appropriate model has to be created. The generation of a Coloured Petri Net (CPN) out of the static and dynamic characteristics of a network configuration was introduced at the FAIM in 2007. The static part or the structure of formation is given by the typology or arrangement of the CC in OPN. Dynamic systems show an additional process structure, which contains the logical relations or the time sequence. The process structure of an enterprise in the context of OPN depends on the predicted resource situation and the degree of capacity utilization of the CC or the way the CC offered its services in the EVCM-specific offer phase. During the modelling, a quasi insulated system was considered typically. However, real systems are open and differ alone in their exchange relations to the surroundings of insulated systems. Due to this, the model only contains a restricted usability regarding a realistic modelling and simulation of material flow of OPN [6]. It applies in general that the simulation represents an approved measure of the preventive and reactive disturbance management for an evaluation of the effectiveness of decisions, e.g. in case of a reconfiguration. The presented measures for a stable operation of OPN are completed by further measures out of the field of research of cybernetic principles in the next chapter.

CYBERNETIC PRINCIPLES

Nowadays, cybernetics is understood as the science of control of dynamic systems of different material quality. Its primary concern is to make general statements about the structure and functional principles of a purposive impact on the behaviour and motion of such systems [21]. It is differed between real and abstract systems. Abstract systems are for example socio-technical systems [22]. The value-added chain is also characterised as a socio-technical system. The application of the cybernetic approach seems to be advantageous, as a broad spectrum of measures for the suppression of disturbances has been developed with it, whose capability is proven. Cybernetics in enterprises offers opportunities to influence the subsiding company processes in the sense of a feedback control, under appraisal of external exposures as disturbance variables [9]. The following conception of a disturbance-sturdy network operation herein contains a relatively high level of abstraction, so that an important precondition for the application of cybernetics is given [20]. In the following, presented measures are allocated to factors, which are relevant for competition: dynamics, complexity and time.

DYNAMICS

OPN contain a changeable as well as a changing formation- and process structure, typically for dynamic systems. At this point, the structure is imagined as an arrangement of elements in space. The structure of OPN, compared to the strategic networks, is subject to a distinct temporal variance. In all abstract systems, also in OPN, the structure can be influenced or arranged consciously. If the term structure is used in a way that it is understood as a pattern of a special arrangement or sequence of events under a temporal aspect, it is defined as a process structure in literature [14]. A sequence of interrelated activities is summarized as process and the existence of a process structure is deviated out of it. The process structure also changes over the time. This in general applies for enterprises. Regarding primary value-added activities (e.g. manufacturing of requested products), OPN contain a process structure. Through changes of the formation- and process structure, the behaviour of the system changes additionally. In summary it can be said that a structure is not only defined as an arrangement of elements in space, but also as a temporal arrangement of its activities [22]. In time-variant systems, it is important to permanently collect the performance of control and to ensure its quality by means of selected interferences. If this context is
assigned to the disturbance management in OPN, the degree of efficiency of the suppression of disturbances has to be checked and, if needed, the concept of interference suppression needs to be corrected. The already presented measures of the key figures delivery reliability or modeling and simulation can be applied supportively in this context. For the collection of control performance, the already demonstrated monitoring concept serves as the basis.

**Complexity**

The value added chain belongs to dynamic and complex systems. The suppression of disturbances is complex in such systems. ASHBY pointed out that “Only Variety can destroy variety”[14]. Therefore, approaches for the transmission of cybernetic principles to production, in many cases demonstrate a simplification and decentralization of system structures. There is an intensified concentration on product variants or the simplification of the manufacturing organization. The complexity of a system, among others, depends on the degree of centralization of a structure. It is therefore differed between centrally and locally organized systems. In case of a local process organization, the behaviour of the performing subsystems is considered as “black box” accordingly. In OPN, exactly this approach is pursued. The production is organized locally. Considered abstractly, the production net consists of performing subsystems and the flows of material and information which run between them. The enterprises obtain a production order during the network operation. Production occurs independently, without controlling interventions of the network management. In local systems, the range of possible reactions on disturbances is reduced to disturbances, which are specific for subsystems. For each disturbance variable, an alternative action is allocated to the local controller [14]. In OPN, the network management has no influence on locally occurred disturbances. The enterprises have to resolve internal disturbances on their own. A preventive elimination of causes is recommended.

In this context, Hilli recommends a disturbance feedforward directly at the place where the disturbance occurs. Thereby, effects of disturbances shall be avoided. The Kaizen-Concept [15] provides the formal frame. For example, directly concerned employees at the place of disturbance have to be included in the suppression of disturbances [9]. In this respect, a more efficient suppression compared to the suppression in central control loop structures, will arise. In case of locally arising disturbances, whose effects influence the subsequent production steps negatively, the already presented monitoring concept will be applied. By means of allocated WfMS-entities, deviations can be identified and reported in time.

**Time**

Besides variety, the dead time signifies a further parameter for system stability and the performance of control. The dead time describes the delay of the control path between the variance of the correcting variable of the controller and the coming into force of this variance within this control variable. The stability is endangered, if the control path reacts less strongly to the variance of the correcting variable with its output variable than the variance adjusts. Concerning the value-added process, the consequence in this context is that the processing time represents an essential factor for the dead time. The shorter the processing time of primary activities, the faster does a feedback and therefore a reaction to disturbances occur. The system more quickly aspires to a new equilibrium state. At this point, a possible inconsistency to the approach of the delivery buffer, which has been discussed already, can be observed. Where stocks provided for a decoupling against perturbations, a shorter processing time shall be achieved, e.g. by means of a changed process structure. This would be consistent with a flow concept, which says that buffers represent a waste of time and therefore have to be eliminated. On the one hand, delivery buffers serve as a measure of reliability for an adherence of delivery dates and on the other hand, they influence the reactivity of the control variable to deviations of the controller negatively. It is the task of the network management or the disturbance management to resolve such conflicts.

**Feedback vs. Feedforward Control**

For a target-oriented engagement, the principles of feedback and feedforward control are to be distinguished in the language use of cybernetics as basic types of control. In case of disturbances, a target-conform reaction shall be actuated to cause the desired state [22]. Most of the processes within the added value are influenced according to the principle of feedback control. This also applies for the reaction on deviations within the functional operation of the network. The deviations are recognized by the distributed WfMS and communicated to the disturbance management, if needed [5] [8]. Whereas the feedforward control only determines processes and measures previously and is therefore able to compensate common disturbances in advance, the feedback control affects the process permanently, as relevant data are measured and supplied to a controller, which in case of deviations to the expected process, initiates appropriate corrective measures. Therefore, the controller cannot take effect before the incidence of disturbances. But in case of a frequent measurement and the relevant strategy, disturbances can be balanced rapidly.
and cost-efficient by the controller. This shall provide for a stabilization of the system’s behaviour also in case of a high disturbance effect and unknown cause-and-effect relations. At this point, a critical comment has to be made concerning references [5] and [8], which postulate the feedback control as control strategy. It can be noticed on one side that causes for disturbances in subsystems or enterprises will often remain unknown to the network management. The reason for this lies in the temporary formation- and process structure of the OPN. This structure changes with each new order. On the other hand, it applies for each network type that through manufacturing of small order quantities, a high process frequency, which would contribute to an optimal control loop parameterization, cannot be guaranteed. In summary it has to be noted that feedback control offers advantages compared to feedforward control at the first glance (unknown relation of cause-and-effect), but the realization of a stable operation of the network according to the feedback control principle is subject to conditions which have not been solved sufficiently so far. The feedforward control principle, however, requires an active treatment with the process. In strategy-oriented networks with fixed dependencies between the producer and his suppliers, an active treatment with the production process can occur. The JIT-suppliers for example have to deliver accessories synchronous to the manufacturing process of the producer. It is important in this matter that the focal enterprise can influence the manufacturing processes of his suppliers, e.g. to optimize the processing times corporately.

A further difference between feedback and feedforward control lies in the runtime performance of the systems. Feedforward control operates immediately and free of vibration, whereas the feedback control needs time (dead time) for an activation of counteractive measures in case of deviations, which makes an absolute stability of the controlled system impossible regarding the command variable. For a decrease of such dead times, it can be tried to improve the flow of information (see WIMS-concept).

SUMMARY

Eventually, the problems during the transmission of cybernetic principles shall be summarized to the value-added process. Deviations are value-free within cybernetics, i.e. positive deviations are suppressed to the same extent than negative ones. Deviations have to be evaluated, wherefore the disturbance management in OPN is responsible. For the evaluation the simulation of the flow of material can be applied. Afterwards, it can be decided on the application of an appropriate control strategy. A general problem of cybernetics is that it does not differ enough between the cause and effect of disturbances. Cybernetics is effect-oriented to one side, which implicates advantages during feedback control but on the other hand, as already mentioned, complicates the correct adjustment of the control variable. A successful control loop system can hold the process stable but is, above all, unable to optimize the process.

REFERENCES

Supply Chain Management Constraints: Telecom Industry Case Study

Fatiha NAOUI*

Department of Management Sciences (NIMEC)
Enterprises Administration Institute, Caen University. 3, Claude Bloch Street, 14075 Caen / La Rochelle Business School - 102, Coureilles Street Les Minimes - 17024 La Rochelle Cedex 1.

ABSTRACT

This paper treats of the Supply Chain Management as an organizing concept and also discusses of its’ principal implementation constraints. This study is based on semi-directive interviews with telecom industry companies’ personnel members and direct observation. The Telecom Industry example is employed to illustrate the actor’s perception. The objective is to emerge the SCM constraints dimensions. In this study, several levels of SCM implementation constraints are identified. The generic level constraints, the trade level constraints, the information systems level constraints and the subcontracting level constraints. These dimensions allow drawing up framework actions to improve the SCM concept performance within the telecom industry.

Key words: Supply Chain Management, Implementation, Constraints, Telecom Industry Case Study.

1. INTRODUCTION

The SCM domain illustrates perfectly the type of concern in terms of identification and risk management. Indeed, the development of integrated organizational structures and the transverse practices of manager based on process management, interesting situations to be studied. The imperative of customer performance type ECR (Efficient Customer Response) can come true only within organizations endowed with indicators performance versus vulnerability indicators susceptible to estimate clearly the value chain degree of reliability. A global management of the SCM risks is imperative.

The changes of behaviour of the worldwide market require an agile answer from the company and its partners in the supply chain. A true competitive advantage is obtained when the organization can uniformly reply to customer requirements with more precision and reactivity. The past was characterized by standard products, mass-produced for expected demands. Currently, the customers require more customized products in small quantities with a degree higher of uncertainty [1]. The organizations, which in the past felt were protected from foreign competition at low prices, must create value for their customers at a lower price [1]. The customers are the first concern of any firm, which seeks to be competitive. From now on, the logistics of operation with the customer forms an integral part of the customer’s needs and their perception of the firm’s performance [2].

Several questions are open. In particular, it is necessary to define what the concept of SCM covers and under which angles the performance of the company can be evaluated from this point of view. From there rises the central question, which we are trying to answer: Which are the SCM implementation constraints? This paper provides answers to this question. Indeed, this document is articulated around five-sections. The three first ones are devoted to the theoretical delimitation of the supply chain management concept and the SCM constraints. Then, in the methodology section we outline how analysis of the case studies was conducted. The results highlight a number of consistent themes across all concepts which are discussed and some propositions are provided. The conclusion argues that future research should increasingly embrace the multi-criterion and multidimensional approach to the analysis of SCM implementation constraints.

* Corresponding author : Tel. : (0033) 231 56 66 77; Fax: (0033) 2 31 56 65 65; E-mail: fatiha.naoui@unicaen.fr naouif@esc-larochelle.fr
2. SUPPLY CHAIN MANAGEMENT: AS AN ORGANIZING CONCEPT

Many definitions of the SCM are presented in the literature. Several academic research studies highlight the role-played by logistics, judged as being at the origin of the step SCM [10]. Reference [11] proposed SCM as systemic, strategic coordination and the tactical management of the actions within the departments of a particular organization, as well as businesses carried out inside the chain of provisioning of the organizations as a whole.

2.1. THE STRATEGIC AND INTEGRATING ROLE OF THE SUPPLY CHAIN

The network of the organizations which requires, through bonds upstream and downstream and in various processes and activities to produce value in the products and services held in the hands of the ultimate customer [12]. According to the author, it is initially by the "4p" model that the emergence of SCM is explained. The product, the price and the promotion, while the fourth element, described by the right product, in the right place, the right time, concerns a logistic dynamic. Reference [13] Added to this approach three competitive advantage factors: [Reliability, Responsiveness, Relationship]. The reliability of an organization is dependent on the need for guaranteeing a delivery complete and on time; then, the sensitivity (upon request) evaluates the capacity to answer within the shortest possible time, with the largest flexibility; finally, the relationship factor stresses the importance of partnerships in the implementation of continuous improvement as regards quality, innovation, reduction of costs and adjustment of the delivery logs. In the same way, every company sensitive must proceed to a major reorientation of its management system [14]. Thus, it must modify its organizational diagram according to four points: - To change from a system in terms of functions to logic in terms of process. This means that the company must regard the horizontal character of the structure as a unit of inter-functional processes based on the requirements of the customer; - To change from a profit concept to a performance concept. This point underlines the obligation to provide the financial and non-financial indicators; - To change from products management to customer management: because the customers’ satisfaction must be the ultimate objective of any commercial organization, it is imperative that the structures of management and the systems of measurement are also reflected [14].

2.2. THE WORLD CLASS LOGISTICS MODEL

The model provides a nomenclature in four key dimensions. Initially, strategic positioning selects the strategic and structural approaches, which guide the logistic operations. At the second level, integration establishes what is advisable to do and how to carry it out with reactivity. At the third level, the agility is intended like the aptitude to obtain and maintain competitiveness as with obtaining the confidence of the customer. The criteria of anticipation, adaptability and flexibility attempt to answer this field of performance. It is a question of remaining vigilant of the customers’ unexpected requests, while being able to decrease the response time to the exceptional requests. This highlights the adaptability of the structure at all unexpected circumstances. Finally, at the forth level, the performance measurement evaluates the internal and external logistic chain. The company determines choices of static and dynamic internal indicators enabling it to refine the process supply chain evaluation by benchmarking techniques, which enrich and diversify the evaluation modes.

2.3. SUPPLY CHAIN 2000 STRUCTURE

As in [15] this structure proposes a sequential step, since the reasoning proposes the relations and methodologies to apply to arrive step by step to coordination between the individuals and the organizations involved in a SCM step. In the other words, from a methodological point of view the sample must seek to gather internal and external actors who revolve around the same finality of product or a service design. This structure seeks to clarify the comprehension of SCM as a step of strategic management. Other more recent works specify certain features of SCM such as the role of information systems, confidence and power in the inter-organisational exchanges.

2.4. SUPPLY CHAIN MANAGEMENT TYPOLOGIES:

Reference [16] proposed 4 SCM typologies: within the organization; Dyadic relation; several external chains organizations for a vision a general direction. If SCM is appreciated, first of all, within an organization, Harland thinks seriously of applying an analysis in terms of inter-organisational relations, which are involved to connect two units between them.
3. THE SCM IMPLEMENTATION CONSTRAINTS: A COMPLEX STEP

3.1. NEW MODELS OF ORGANIZATION: TOWARDS NEW CONSTRAINTS

The examples can be given to illustrate the vulnerability of the SCM. Example, Toyota was brought to interrupt a factory of production following a fire in the warehouse of one of his suppliers. The cost of this interruption was estimated at 40 million dollars by days (Nelson and al., 1998). This illustration translates the possible consequences dysfunctions of a chain link on the global logistic system. At the same time as these various sources of risk with difficulty predictable nature, [17] underline that numerous factors during the last decade stressed the sources of risk within the supply chain. These factors are contingent to the logistics and production system. The companies’ praiseworthy objective directed to the efficiency versus the efficiency leads to have an integrated supply chain thus more and more dependent on the weakest link. Also, the economic globalization and the strategic development of outsourcing increase the complexity degree of the supply chain. The network henceforth consisting of side and horizontal links, feedback [18], the property lack and the development of the slowness lead to weaken the chain.

Finally, our analysis is not exhaustive, the suppliers’ number reduction tendency, make without a preliminary rigorous selection, constitutes an important source of risk. It’s the same if their reproduction drives to mask and to aggravate the failure sources rather than to limit their probability of case.

5. EMPIRICAL PART

5.1. METHODOLOGY

The case study is one of several ways of doing social science research. In general, case studies are the preferred strategy when “how” or why” questions are being posed, when the investigator has little control over events, and when the focus is on a contemporary phenomenon within some real-life context. In order to discover certain explanatory dimensions of the SCM implementation constraints, the methodology was qualitative, using semistructured interviews based on 15 respondents. This method retained constitutes one of the collection modes most answered in the field in management science of research [19].

Table 1: The Alpha respondent’s statute and interview duration

<table>
<thead>
<tr>
<th>RESPONDENTS STATUTE</th>
<th>INTERVIEW DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution Data Processing Responsible</td>
<td>9.30 at 11.40 H</td>
</tr>
<tr>
<td>Operational Responsible</td>
<td>14.00 at 16.30 H</td>
</tr>
<tr>
<td>Reporting Responsible</td>
<td>14.00 at 16.00 H</td>
</tr>
<tr>
<td>Supply Chain Manager</td>
<td>17.00 at 20.00 H</td>
</tr>
<tr>
<td>Planning Responsible</td>
<td>14.00 at 16.00 H</td>
</tr>
<tr>
<td>Quality Responsible</td>
<td>9.30 at 11.40 H</td>
</tr>
<tr>
<td>SAP Manager</td>
<td>10.00 at 12.00 H</td>
</tr>
<tr>
<td>Ingineer</td>
<td>16. at 18.30 H</td>
</tr>
<tr>
<td>Customer Relationship Manager</td>
<td>10. at 11.30 H</td>
</tr>
<tr>
<td>SAP Ingineer</td>
<td>17. at 19.00 H</td>
</tr>
</tbody>
</table>

First, the methodological aspects of research are presented. Second, we present the results of our analyses. Methodology is approached through three points. Firstly, we introduce the Alpha company choice, then the discussion and protocol used, and finally the principal elements relating to the content analysis.
Case Study presentation: Alpha is organized in five activity groups and four geographical areas.

Table 1: The Alpha Geographical Areas and Activities Segment

<table>
<thead>
<tr>
<th>ACTIVITY GROUPS</th>
<th>ACTIVITY SEGMENTS ROLE</th>
<th>GEOGRAPHICAL AREAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator Segment :</td>
<td>Operators: Suppliers service need</td>
<td>Europe and North</td>
</tr>
<tr>
<td>Mobile, Fixed, Convergence</td>
<td>Services: Conceives, Deploys, Manage</td>
<td>Europe and south</td>
</tr>
<tr>
<td>Services Segment</td>
<td>Network Maintenance.</td>
<td>North America</td>
</tr>
<tr>
<td>Companies Segment</td>
<td>Companies: Meeting the companies needs</td>
<td>Asia-Pacific</td>
</tr>
</tbody>
</table>

Research Participants: Before the professional data collection, I could have discussed my research with several academic professors. The staff members questioned at the company was confronted with many corporate during their professional experience within Alpha and they are interested in the SCM step. I asked Alpha company to choose staff who were knowledgeable about SCM. They suggested other staff as well who could be useful in my research.

Discussion and Protocol: For interview technical choice we use semistructured interviews. They were held starting from a flexible interview guide, was prepared in advance. Several themes are approached. The use of a tape recorder is necessary in order to allow the interviews re-transcription. The interlocutors (information system team, Quality, SCM, Service Support, Operational and Reporting department) were not reticent of the use of the tape recorder and analyzed them with the 48 hours following the interviews. The interviews lasted one to two hours. All the interviews started with a general question. We let our respondents’ approach the questions freely, while including them to go further in potentially interesting points [20]. The objective is to reveal implementation constraints within the Alpha SCM dynamics.

Data Collection and Analysis: The Alpha data collection was completed in February and Mars 2008. I also collected throughout the study various documents allowing me to increase my analyses. They were mainly the studies of consultants, newsletters, and annual reports. This step is inspired by Miles and Huberman typological analysis [21]. To condense me qualitative data, we coded the interviews. The analysis of interviews was done starting from a categorical analysis set of themes on the sense units present in the interview guide. The categorical analysis is a powerful device of data condensation, whose fundamental principal is similar objects regrouped under a common title or class [22].

5.2. RESULTS

In this section, I examine the research results of SCM implementation constraints, the dimensions were already selected. In answer to the question: What we are talking about, when we talk about SCM implementation constraints? Several constraints arise from the respondent’s comments.

Alpha SCM implementation constraints finding: a complete overlap: In the Alpha case, several levels of SCM implementation constraints show through clearly. Indeed, the appreciation of the respondents questioned appreciations are slightly same. According to the internal actors of the organization, The Alpha SCM implementation encounters the several levels constraints. The generic level constraints, the trade level constraints, the information systems level constraints and the subcontracting level constraints could be identified (cf. following table or stamp).
Table 1. The SCM implementation constraints: a complete overlap

<table>
<thead>
<tr>
<th>CONSTRAINTS</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERIC LEVEL</td>
<td>Organization constraints:</td>
</tr>
<tr>
<td></td>
<td>- Company perimeters change adaptation constraints.</td>
</tr>
<tr>
<td></td>
<td>- Export lawful constraints.</td>
</tr>
<tr>
<td></td>
<td>- Objectives of the management constraints.</td>
</tr>
<tr>
<td></td>
<td>- Resources and the individual constraints.</td>
</tr>
<tr>
<td></td>
<td>- Partners’ constraints become Alpha constraints</td>
</tr>
<tr>
<td>TARES LEVEL</td>
<td>- To find the good compromise.</td>
</tr>
<tr>
<td></td>
<td>- To adapt to the market trends.</td>
</tr>
<tr>
<td></td>
<td>- To speak common language.</td>
</tr>
<tr>
<td></td>
<td>- To manage the service contracts.</td>
</tr>
<tr>
<td></td>
<td>- To automate of the end to end.</td>
</tr>
<tr>
<td></td>
<td>- Temporal and Financial constraints.</td>
</tr>
<tr>
<td>SUBCONTRACTING LEVEL</td>
<td>- To treat the visibility lack.</td>
</tr>
<tr>
<td></td>
<td>- To answer the strong and unforeseen orders.</td>
</tr>
</tbody>
</table>

Source: Table realized from the interviews extracts

The organization constraints level: At the generic level the group meets limits of adaptation. That is related to the company perimeter change. In consequence, in internal, the employees and in particular the data processing specialists seem to have an extra work which is translated by difficulties of respecting time delivery engagements. An organization like Alpha, with objectives of SCM profitability and efficiency, at the time of export in Asia for example, requires the knowledge individual resources and competences. These elements can condition the resources structuring.

All the chain links are concerned and the constraints of the ones influence automatically the other partners of the chain. There is thus an interdependence of the supplier-customer constraints which appears clearly in the Alpha case.

Interviews Extracts:
“I am a Frensh company which wants to export in China. It is already a difficulty” (Reporting Responsible).
“The change of the company is an activity which costs much energy” (Reporting Responsible).
“In terms of individuals we suffer but we are to pay for that. It is necessary to spend every night. It is really the Business Drogen” (Engineer SAP).
“The constraints of my customers and suppliers become my constraints” (Responsible Operations).

 trades constraints level: At Alpha the level trades constraints seem to be complex and in particular on the information systems level. Indeed, a compromise between the business people and the information systems people seem to be necessary. The information system must be evolutionary to satisfy the internal and external customer. The evolutions of the hardware towards the software for example, imply the language difficulties reduction. Indeed, the language addressed to the customer and to the manufacturer is not the same one. Nevertheless, in the SCM an information system is in perpetual evolution and the “end to end” automation seems to be difficult. That is explained by the fast evolution of the market trends compared to the information systems and by the generated cost.

Interviews Extracts:
“People of the ground are sometimes obliged to make manually certain operations” (Information System Engineer).
“It is always for yesterday, it does not have there money, give the best customer service. Be unrealistic and ask me the impossible on” (SAP Responsible).
“When we sold the products it was difficult for the customer to speech of charts. When we arrive at the manufacture level, we passes from function user to a chart and is not at all the same thing” (Business Responsible).
Supply Chain Management Constraints: Telecom Industry Case Study

Subcontracting constraints level: It is significant to specify that Alpha is on a rather pointed market where the forecast and planning are very significant.

Forecasts errors: The subcontracting is the principal constraint which emerges in an obvious way. Indeed, Alpha does not manufacture almost anything. It launches production schedules in the subcontractor. Thus, the group loses on the flexibility level to answer the urgent and unforeseen requests. The errors of the forecasts remain, in spite of the twelve months of visibility given by Alpha to the suppliers.

<table>
<thead>
<tr>
<th>Interviews Extracts:</th>
</tr>
</thead>
<tbody>
<tr>
<td>“A great constraint for us when it’s a question to answer to the significant volume” (Operations Responsible).</td>
</tr>
<tr>
<td>“Sometimes the reality is different from what we imagined. The orders cannot decide day at the following day because our suppliers have the other suppliers” (Responsible Reporting).</td>
</tr>
</tbody>
</table>

The end of the quarter problems: Indeed, the customers market is organized per quarter and at the end of each quarter these budgets must be consumed. Finally, the addition of sale is difficult to manage by Alpha because Alpha to undergo the customers’ budgetary constraints effects.

<table>
<thead>
<tr>
<th>Interviews Extracts:</th>
</tr>
</thead>
<tbody>
<tr>
<td>“At the end of the quarter, this large constraint influence on the Alpha performance” (Responsible Reporting).</td>
</tr>
<tr>
<td>“For example, in Peaceful Asia Areas, Alpha requires of the customer to order before the end of quarter. In this case the customer will be sure delivered in time” (Operations Responsible).</td>
</tr>
</tbody>
</table>

An exhaustive list of SCM implementation constraint is not provided but these proposals would have helped to improve SCM performance to reach the following Business objectives: Improve communication with the customer (operational reporting for customer follow-up); Improve operational reporting (Internal KPIs, Quality report); Improve invoice control and KPIs follow-up ; Improve cost management (End-to-end margin analysis per customer contract).

6. SUMMARY

This paper has the aim of leading a qualitative analysis next to Alpha internal actors. The summary of the results obtained can be articulated by two principal objectives: - the prime objective relates to the apprehension of the SCM implementation constraints by various internal actors questioned. The given answers and remarks confirm that a clear definition and supplement stakes attached to SCM implementation constraints constitutes an essential antecedent. Indeed, this type of project led to a significant change within an organization, which requires a continuous effort of clarification - the second objective relates to the interpretation and the representation with the various actors of the topics suggested. The results obtained are of two types. On the one hand, the analysis of the SCM implementation constraints according to multi-criterion and multidimensional approach is perceived. At the end of this paper, several strong ideas emerged. The successful installation of SCM rests on the aptitude to evolve the behaviours of the organization on all the levels. The organization is reconsidered in-house and in its relations with the customers, the suppliers and the people receiving benefits. The company must be proactive before the cultural change that SCM implies and accept the decompartmentalization. It must share its resources and its information with its partners, under penalty of desynchronising the rest of the chain and to create bottlenecks. All the functions will be concerned: related purchase, commercial, production, logistics and functions, data processing, finance (which will provide indicators to measure the performance and the effectiveness) and human resources which will become a true added value. SCM is thus the search for an excellent total performance in a chain made up of companies, independent but bound by a common objective: the satisfaction of the final customer.

REFERENCES


Supply Chain Management Constraints: Telecom Industry Case Study

How Quality Management could improve the Supply Chain performance of SMES

Paul Eric Dossou*, Philip Mitchell

CIM Department
ICAM Group, IST-Vendée
La Roche-Sur-Yon, 85000, France

ABSTRACT

GRAI Quality is one of the modules of the GRAI methodological tree. It allows to implement and improve knowledge management in general, and more especially quality management in enterprises. Two modules called GRAISUC for choosing and implementing Supply Chain Management (SCM) Tools in SME’s and GRAIQUAL for implementing, managing and improving quality in SME’s are being developed. After presenting briefly the concepts of GRAI Quality and the architectures of GRAIQUAL and GRAISUC, we will focus our attention on how GRAIQUAL, will be able to define quality parameters throughout the enterprise supply chain, choose the right tools according to the enterprise in question, manage the approach of implementation and improve the supply chain performance in terms of quality, cost and lead time. Then we will show the interaction between the choice and the implementation of a SCM tool managed by GRAISUC and the performance improvement in terms of quality management guided by GRAIQUAL. We will illustrate the different concepts and architectures used by presenting real enterprise applications in SME’s.

1. INTRODUCTION

The actual difficulties of enterprises due to the new economic crisis with its financial constraints in the context of increasing globalisation impose on SMEs a high degree of organisation for survival. It has become essential to have a high-performance supply chain. The GRAI methodology [2] allows to reorganise the enterprise, to choose and implement software solutions, to manage knowledge, to define and realise industrial strategy, and to manage the global performance of the enterprise. The supply chain management tools are an interesting answer for SMEs [5].

GRAI quality is used for organising the implementation of a global quality approach in the enterprise as well as quality management (product quality, process quality, and system quality). The implementation of a global quality approach [7] is a pertinent answer to SME difficulties, whose main objective is the improvement of their global performance. To achieve this, the defined approach proposes a rationalisation of the process for choosing and implementing global quality policy for the enterprise concerned [6]. The use of this tool and the different concepts allows to improve the performance of the supply chain. Indeed, we are convinced that the main performance criteria of the enterprise (quality, cost and lead time) used for optimising supply chain performance are interrelated [1]. So, the optimisation of the supply chain becomes the search for a compromise between these three criteria.

In this paper, we will show that the implementation and the use of a global quality approach in an enterprise has a real influence on its supply chain in terms of optimisation. After the presentation of the concepts of GRAIQUAL we will show how a quality system could change the whole supply chain efficiency, then we will illustrate the concepts with a case study.
2. GRAIQ QUALITY APPROACH AND GRAIQQUAL ARCHITECTURE

GRAI Quality is based on a GIM (GRAI Integrated Methodology) approach. So it starts with a modelling of the existing situation, in order to obtain an ‘AS IS’ model (figure 1). The objectives are to take a photo of the enterprise in question, showing the state of the enterprise at that moment of time. It will be used for detecting the inconsistencies in the enterprise. Some strong points and points to improve are deduced from these inconsistencies. Then, future models are established (TO BE). These models are obtained after the design phase of GIM (GRAI Integrated Methodology) approach.

In the GRAI QUALITY approach, the ‘SHOULD BE’ is the « vision » of the future system and is considered as the main point in the enterprise’s development. The ‘SHOULD BE’ elaboration consists of the definition of a long term Business Plan including the strategic objectives. The ‘AS IS’ and ‘TO BE’ models have the same structure (decisional, functional, physical, informational, process). The ‘TO BE’ models are the result of a combination the ambition of the enterprise and constraints related to the existing system and real possibilities imposed by the environment and the market.

From the obtained future model (TO BE) and the ‘SHOULD BE’ integrating the strategic objectives of the enterprise, GRAI Quality allows to develop a new approach to quality implementation. We can notice a similarity with GRAIPROGI (GRAISUC used for choosing and implementing a supply chain management tool) during the first phases. The difference concerns the following steps, that is to say the exploitation of the results obtained during the first phases. Not only will they be used for defining functionalities of a handbook in order to choose a software tool and to propose an efficient action plan, but also they will allow to define a quality policy and the implementation of an action plan.

The definition of action plans related to quality and their implementation act both on the physical structure of the enterprise and on the process as well as the manufactured products. Thus the whole supply chain benefits from the improvements obtained.

For the definition and the application of a quality policy in the enterprise, we used the GRAIQQUAL tool. The concepts of this tool are presented in [6]. They are a continuity of the concepts elaborated for GRAIXpert the hybrid expert system [3], [10]. Two possibilities of knowledge refinement were defined : the refinement of the knowledge by adding a distinct knowledge and the refinement of the knowledge by combining existing knowledge [8], [9].

Obviously these two possibilities of knowledge refinement depend on the nature and mode of representation of the knowledge acquired. As far as the acquisition of enterprise knowledge is concerned we propose the use of three modes of knowledge recognition: reference models, case studies and rules. For the reference models, the knowledge to add will be a refinement of the existing knowledge. For the case studies the objective is to continue to enrich the GRAIXPERT data base. This means the exclusive addition of distinct will be applied.
How Quality Management could improve the Supply Chain performance of SMES, FAIM2009

The concept of reference model allows with GRAIXPERT to define according to the domain the optimum realizable in terms of the quality approach and the implementation of certification. For example, this could be the management of the process and all the necessary stages to obtain ISO 9000 version 2000 with all the implementation phase and follow up.

In order to manage all the GRAI Quality approach and the design of the ‘TO BE’ models the problem resolution method developed for the expert system GRAIXPERT is used. It is based on several reasoning mechanisms: CBR (Case Based Reasoning), Decomposition, Transformation and direct correspondence [3]. The architecture of the new supporting tool GRAIQUAL is also presented. In this architecture we have a Management Module (MM), a Work Base (WB), a Transfer Interface (TI), Improvements Management Module (IMM) and a Knowledge Base Module (KBM).

The Management Module is used for organising the different interactions of the tool with the expert system (GRAIXPERT) and the kernel of e-Magim. The Work Base is being elaborated for managing and capitalising knowledge about the studied case. It’s the space used for improving the enterprise especially the quality system of the enterprise. The Transfer Interface is used for putting the new case in GRAIXPERT in order to improve its Cases Base. The reference model elaborated for each enterprise domain will be improved by the acquisition of new models in GRAIXPERT. The Improvements Management Module is also being developed for managing the different quality action plans of the enterprise. It contains different quality tools. A Knowledge Base Module is being elaborated for containing the rules according to quality certifications.

In the IMM we can notice for example tools like SPC (Statistical Process Control), Poka-Yoke, QFD (Quality Function Deployment), PDCA (Plan, Do, Check, Act), Hoshin, Kanban, 5S and Lean Manufacturing.

In the KBM we can also see norms such as HACCP, BRC or IFS for the agribusiness sector. The HACCP (Hazard Analysis Critical Control Point) consists in identifying and evaluating the risks associated with various stages of the production of edible products and to define the necessary means to master them.

It must be used as an approach which is reasoned, organized and systematic to give total confidence in the product in terms of health and safety demands. Other norms are also present such as SQA (Suppliers Quality Assurance) management, the labels, norms NF, CE, ISO TS 16949 for the automotive sector, AOC for wine, etc.

Once defined the architecture allows to bring to the company all the expertise in the field of quality, continuous improvement tools and certification procedure. How the Quality system works is illustrated in figure 2.

One of the most important factors of the model developed is equally the opportunity by means of a comparative study to show to the decision makers of the company the interest in setting up a procedure for quality, continuous improvement, or certification by carrying out a cost study in non quality. Investment in quality management is always profitable in the medium term, even if decision makers often demand instant results. The economic and financial arguments to convince them are made more credible by the use of this module.
3. THE IMPACT OF QUALITY MANAGEMENT ON THE SC PERFORMANCE

We can define \( E \) as a vector space corresponding to the supply chain of an enterprise. We associate to this set two operations: the sum (+) corresponding in our case to the sum of two elements of \( E \), and the product (*) representing the multiplication of a real number by an element of \( E \), or the product of two elements of \( E \). We can define a \( \mathbb{R} \) commutative body associated to an object \( O \) (here the supply chain). We can define by the reasoning of decomposition, the theory of systems and the theory of hierarchic systems, some sub-objects \( O_k \) associated to each part of the supply chain. We obtain:

\[
O = \bigcup_{k=1}^{n} O_k
\]

with \( n \) equal to the number of sub-objects. These sub-objects correspond to components of the supply chain. We can mention the different processes but also the controls and means allowing to manage this chain.

We can define \( L_1 \) an endomorphism of \( E \) and \( u \) a vector of \( E \). We also define \( u_k \) the vector of \( E \) associated to the sub-object \( O_k \), \( u_k \) being a basic vector of \( L_1 \). If \( \lambda_k \) is the basic value associated to \( O_k \), we can deduce that:

\[
L_1(u_k) = \lambda_k * u_k
\]

We obtain the following linear combination by extending this relation to the object \( O \):

\[
L_1(u) = \sum_{k=1}^{n} \lambda_k * u_k.
\]

The endomorphism \( L_1 \) broken down from the vector \( u \) of \( E \) associated to the object \( O \), into basic vectors \( u_k \) associated to each sub-object \( O_k \) and object \( O \) into sub-objects \( O_k \). Furthermore, the choice of basic vectors as basis of decomposition introduces conformity in the realisation of a good design [6].

![Figure 3: Supply chain processes](image)

Each supply chain is improved by optimising the main performance criteria such as quality, cost, lead time. Let us focus our attention on quality. Each vector \( u_k \) corresponding to a given sub-object will be composed of vectors \( q_1 \), \( q_2 \), \( q_3 \), \( q_4 \), associated respectively to the main aspects of quality. The vector \( q_i \) represents Quality of suppliers, \( q_2 \) Quality of products, \( q_3 \) Quality of process, and \( q_4 \) global Quality of the system. These vectors are defined for each sub-object and indicate the global state of the sub-object according to the performance criterion Quality. We obtain:

\[
L_1(u_k) = \sum_{i=1}^{4} (\alpha_i * q_i)
\]
How Quality Management could improve the Supply Chain performance of SMES, FAIM2009

\[ L(u) = \sum_{i=1}^{n} (A_i \ast (\sum_{j=1}^{k} \alpha_{ij} \ast q_i)) \]

The following step is the design and local optimisation. The global optimisation objectives are detailed in local attainable objectives for a sub-object. An optimisation of criteria is done for each object. The coefficients \( \alpha_{ij} \) associated to vectors defining \( u_k \) are optimised by using the reference models defined in GRAIQUAL. The basic values \( \lambda_{ik} \) associated to each sub-object are implicitly optimised and by deduction the sub-object. We transform by successive iterations the sub-object \( O_k \) into designed sub-object \( O_k^c \).

Indeed, we define a vector space \( F \) associated to the designed object \( O^c \). It has the same dimension as \( E \). Let us also define \( L_2 \) a linear application from \( E \) to \( F \) which transforms each basic vector \( u_k \) associated to the object \( O_k \) into a vector \( v_k \) associated to a designed sub-object \( O_k^c \) as follow :

\[ v_k = L_2(u_k) = \delta_k \ast u_k \]

\[ L_2(v) = \sum_{k=1}^{n} \sigma_k \ast v_k \]

The validation of the optimisation of all the sub-objects implies a re-composition stage. It is clear that the sum of local optima is not necessarily the global optimum. In addition to the sum, this phase also guarantees coherence between all the defined optima, according to the existing reference models. The design solution of the object \( O \) is obtained by combining different partial solutions obtained for the sub-objects \( O_k \) and by keeping the coherence of the set. We obtain :

\[ O^c = \bigcup_{k=1}^{n} O_k^c. \]

We can easily notice in the structure formalised above that GRAIQUAL tool is used for managing different aspects of quality. It allows to analyse the quality approach (if it already exists) of an enterprise, and to propose an improvement process, action plan in conformity with reference models along with existing quality tools in GRAIQUAL. The defined process of quality acts on each part of the supply chain.

Figure4: factors of Quality
We can consider that for radically improving the performance of the supply chain, we need to carry out the same transformations based on the other performance criteria. In reality, the definition of a global optimum in terms of quality is not separable from an improvement to cost and lead time. Indeed, the elimination of defects in the manufacturing of a product for example, leads to reduced production and transport costs in order to satisfy customers and avoid penalties or claims due to the bad quality of products. It also improves the manufacturing process and thus reduces lead time. The implementation of a global quality approach greatly influences the performance of the supply chain.

5. EXAMPLE

This example presents the study of a company specialised in industrial joinery and founded in 1999. Its main sectors of activity are aluminium and PVC. The enterprise has a workforce of 130 employees. The company designs and manufactures products in aluminium and PVC. Its turnover is growing steadily and was 13M Euros for the year 2006-2007.

Since 2007, this enterprise has increased its production and customers. Because of different difficulties in terms of lead time, quality, delivery dates, the manager has decided to improve the supply chain of the enterprise in order to be more efficient. For doing that, a study of the enterprise allows to define an action plan for the enterprise. The way to obtain this action plan is outlined below. The application of one action allows to acquire a supply chain management tool in order to facilitate the management of sourcing, purchasing, procurement, production transport and distribution. For the choice and the implementation of this tool, we used GRAI Methodology, and particularly the concepts of GRAIPROGI and GRAISUC [5].

For satisfying the objective of the enterprise, we start the study by using the GRAI methodology. The first step is about the modelling of the enterprise. We take a photo of this enterprise. Then we analyse this photo to detect inconsistencies in the system. So, these inconsistencies are used for defining points to improve in order to obtain a new reorganisation of the company. These are the main points to improve in the enterprise.

- Coordination and management of the different planning,
- Optimisation of the supply chain
- Quality of products,
- Management of quality system,
- Cost of products
- Environment and employees quality of life in the enterprise,
- Innovation,
- Durable development,

In order to satisfy customers and increase their numbers, but also to be efficient and economically viable the enterprise decided to define an ambitious action plan:

- Production system reorganisation within two months,
- Choice and implementation of supply chain management tool within 6 months,
- Implementation of quality approach within one year,
- Preparation and implementation of ISO 9000 and ISO 14001 certifications within 2 years,
- Preparation and implementation of OHSAS 18001 certification within 3 years OHSAS 18001,

The following figure (figure 5) summarises the different domains improved:

This has led the enterprise to reorganise the layout of the factory in order to make the flow more logical and efficient.

An extension of the factory and the dispatch area has been done to increase production and adapt it to the expectations of the market. For this reorganisation, all the enterprise has undergone change due to 5S and Kanban.
A new organisation has been proposed for the procurement process. The enterprise has invested in an automated trans-storage which allows the storage of long pieces of aluminium (6 meters) and their transportation to the production lines according to the moment defined by the different production schedules. This automated storage implies the gain of space.

A global quality approach has been launched on the supply chain. The quality concepts presented above and GRAIQUAL module allows to implement the new enterprise quality policy. The enterprise has been decomposed by using its main processes. We worked with a customer-supplier structure. Each process has one or many suppliers and customers (internal or external to the enterprise). Each process of the chain has been improved by paying particular attention to each vector of quality (Quality of suppliers, Quality of products, Quality of process and Quality of system). Some indicators have been defined relating to each vector in order to measure the evolution of the process according to the chosen criterion. The objective is to follow in short, medium and long terms the progress in quality of the process.

For the procurement process, a SQA approach has been used, with sourcing to find new suppliers, to evaluate suppliers by using a questionnaire, to assess the quality system of the suppliers (is the supplier able to respect our specifications in terms of quality). The automated storage allows to free employees for other activities such as quality control for incoming goods, and control of the supplier’s process. A quality control system by sampling was also implemented upon leaving storage. The global quality system is managed by a quality Procurement manager with coherence of the whole policy of the process and the respect of the objectives. The quality of the product at the level of the procurement has increased by using this new way of working. At the moment of delivery to the main customer of the process (production), the quality of the product is 100%. An unannounced audit is organised every two months by the customer (production) in order to verify the reliability of the procurement process and the conformity with the objectives. The impact on the quality of life of the employees and the safety of the new organisation are clear.

This concept applied to each process of the enterprise is gathered by the global quality manager for a summary and to facilitate coordination of the new management system. We can notice an original idea involving the production process. In addition, we have to respect the environment and build durable development principles. So a temporary storage area was created for reusing aluminium off cuts. Indeed, to make the doors and windows we use as raw material, long aluminium profiles (6meters). The off cut was thrown away, which was damaging both economically and environmentally for the enterprise. The management of the second storage is synchronised with the first one and also with the supply chain management tool in order to choose the right materials for a product.

The process necessary for the certification ISO 9000 is under way. To achieve its objectives the company has set up a formalised quality system in order to conform to the certification norms. It has also identified risks implementing the global quality approach, calculating the risks for the product and has guaranteed the control of these risks.
The process for ISO 14001 is also under way. The industrial activities conform to existing legislation regarding the protection of the environment. The impact on the environment is minimised by using and researching ways to have a reasoned use of resources and protection of the environment in the framework of sustainable growth. To achieve these goals the company is directly responsible for the environmental impact of its activities. It manages its activities accordingly, uses and maintains its production tools (and the supply chain management tool) respecting the environment and informs the workforce as well as its customers of the importance of this policy so that they take into account the preservation of the environment.

The implementation of OHSAS 18001 is under way.

6. CONCLUSION

The three main performance criteria of an enterprise are really related. A focus on quality in SMES allows to improve cost and lead time. Because of the crisis, each SME needs to find ways for reducing cost, respecting lead time and delivering product in conformity with customers specifications. This paper presents GRAI Quality a global module of GRAI Methodology (i.e. methodology used for enterprise modelling) for improving quality in SMES.

This theory is already used to prepare and manage the enterprise during its process of quality, safety, and environment certifications.

A Quality tool GRAIQUAL of this module is presented. This tool in interaction with an expert system GRAIXPERT allows to improve the quality system management of an enterprise (or its creation) and the case base of the expert system in order to refine reference models.

Then the Problem solving method used for improving the supply chain of an SME just by working on quality concepts is presented. An example is given to illustrate how this tool could be used to improve the performance of the enterprise supply chain. This example also presented how the use of GRAIQUAL could be combine with GRASUC one (used for choosing and implementing supply chain management tools in SMES).

REFERENCES

Establishing a framework for building a SC performance system

Alberto de la Calle* and Esther Alvarez

Industrial Organisational Department
University of Deusto
Bilbao, 48007, Spain

ABSTRACT

Innovation has become a key issue for enterprises, sectors, regions and countries. Business strategies are based on the possibility of developing new or improved products, processes and management or marketing methods in order to develop competitive advantages. This shows the need to measure the real effect of innovation on business performance. To measure the impact of innovation on business performance is a difficult task because the relationship between the two is not always clear. Focusing on supply chains instead of enterprises makes it even more challenging. Innovators from across the chains must integrate an increasing number of interdependencies among product development decisions, whilst at the same time addressing customer needs/values, product technical specifications, delivery system capabilities and environmental requirements. The analysis of the innovation impact considering the supply chain as an addition of isolated elements leaves out the effects of collaboration and cooperation between the supply chain members. This paper focuses on the importance of building a common framework that supports the development of a performance evaluation system for the whole SC. A methodology is developed in order to facilitate the comprehension of the benefits of launching collaborative innovation strategies among all the agents of the supply chain. This understanding positively influences the firms behaviour towards the construction of an integrated and profitable supply chain. The proposal takes into account the performance measurement in social, economic and environmental issues.

1. INTRODUCTION

Firms are competing in the context of global markets, rapidly changing technology developments, shorter product life cycles, etc. They are then forced to change or improve their performance continuously in order to meet the customers’ requirements as quickly as possible. In fact, finding business strategies to comply with the latest customer demands or needs is becoming one of the main concerns of many enterprises. However, they are not infallible recipes. The implementation of business strategies looks at diminishing the market uncertainty. The strategic plans are based on qualitative or/and qualitative forecasts covering organisational issues at high managerial as well as operational levels. The inherent condition of a changing environment forces enterprises to periodically review the development of their current strategies. Due to this fact, performance measurement is essential. There is a great amount of information needed to help decision makers to measure the impact and the success or failure of their policies. Furthermore, the evaluation system should be used not only for monitoring performance but also for improving it. Among other possibilities, this paper deals with the innovative collaborative strategies.

It is widely agreed that innovation is crucial for the long-term survival and growth of the firm [1] and it has very important effects on meso and micro economic variables such as productivity growth, competitiveness and employment, although the exact relationship between these variables are not very well known [2].

Innovation has been used to signify either new ideas or practices but it is an idea that could be commercialized. Following the Oslo Manual definition, innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations. According to the EU Green Paper on Innovation, innovation is:

- The renewal and enlargement of the range of products and services and the associated markets.
- The establishment of new methods of production, design, supply and distribution.

* Corresponding author: Tel.: (+34) 944139000; Fax: (+34) 944139101; E-mail: acalle@eside.deusto.es
The introduction of changes in management, organisation design, the working conditions and skills of the workforce.

Following these definitions, innovation could be related to three main areas: product, process and management. Maravelakis [3] outlined the connection between product and process innovation and stated that organisational innovation follows these two dimensions. The innovation process is the combination of activities, such as market research, communication, design, process development, organisational restructuring and so on, which are necessary to develop and support an innovative product.

Innovation establishes its basis on what the customers’ value is and what they are willing to pay for it. Drucker [4] suggested that innovation is based on creativity, systematic and order, and also on unexpected events, but in the end, innovation is based on associating a solution to a need. Companies do not usually innovate by themselves. Some authors explained the interest in collaboration between firms and other organisations for innovation processes [5] [6].

Supply Chain (SC) is an interesting framework for understanding how Collaborative Innovation influences the SC performance. This paper focuses on the importance of building a common framework that supports the development of a performance evaluation system for the whole SC.

The paper is structured into five sections. Section 2 presents the innovation and collaboration in the innovation process. Section 3 explains the proposed framework for building a successful Innovative Collaboration strategy. Section 4 reviews the evaluation methods used for measuring the SC performance. Finally, Section 5 presents brief conclusions.

2. INNOVATION AND COLLABORATION

The development of Collaborative Innovation (CI) is seen as a result of interactive learning processes, i.e. there is often an exchange of knowledge between several organisations involved in innovation processes [7]. CI enables the supply chain entities to rapidly engage and contribute, so that they readily add value given their own position in the SC. CI accomplishes this by:

a) Creating an organisational learning atmosphere; previously, Peter Senge [8] proposed five disciplines for building learning organisations. Learning Organisations are where people continually expand their capacity to get the results they truly desire, where collective aspiration is set free and where people are continually learning how to learn together. This concept and its five disciplines suits perfectly in the SC collaboration for innovation context:

- **System thinking** involves a holistic vision. This would allow individual entities to create a total value model.
- **Personal Mastery**: here each entity is meant to identify where to focus innovation effort to add the most value for stakeholders, while seeing reality objectively.
- **Mental Models**: is meant for learning to reveal the internal pictures of each organisation. This would allow each entity to rigorously scrutinize assumptions and decisions so that others in the chain can consider them and add value by identifying necessary adjustments and/or by embracing them based on clear understanding.
- **Building shared vision** is the ability to share the picture of the future, encouraging genuine commitment and enrolment rather than compliance.
- **Team learning** is the backbone of the learning process, allowing collaborative working teams involved in a supply chain to systematically innovate and improve existing processes / products.

b) Modelling the functional design of a product and its associated processes so that information is accessible to the core entities responsible for product / process innovation and those who support it.

- **Innovation functional requirements**, by analysing these requirements across the supply chain, defining the control logic for the collaborative work, and the required data flow also contained within the functional design.
Establishing a framework for building a SC performance system

- **Validation & Simulation**, where the model can be validated and completeness verified by simulation. A dynamic system based on simulation would better fit the environment, where each function is described as a sub-system and their interactions are highlighted using causal loops.

### 3. Building Sustainable Collaborative Innovation (The SC Framework)

Although collaboration is seen as one of the key elements to innovation, the failure rate of strategic alliances is projected to be as high as 70% [9]. There are many difficulties encountered in achieving a true collaboration between SC members i.e. the difficulties for aligning the internal processes, Information Technology infrastructure, Trust, Organisation design, Competition, Fear of external pressure, the presence of Powerhouses within the organisation and Financial issues [10].

#### 3.1 Factors Affecting the SC

Considering the SC in the centre of the analysis, collaboration for innovation in the SC is directly related to some external and internal factors (see Figure 1). These factors deal with social, environmental and financial issues. From the common understanding of these factors, a solid and sustainable collaborative strategy could be adopted in order to improve the SC performance.

![Figure 1. External and Internal factors that affect SC Innovation performance.](image)

---

**External factors to the innovative SC:**

- **Sector**: The analysis through the 5 forces analysis proposed by Michael Porter [11] determines the ultimate profit potential of a sector. The maturity of the sector should also be taken into account.
- **Government**: Policies, special programs for strengthening innovation, the creation of innovation spaces in which companies can interact with other innovation agents.
- **Innovation systems at macro-meso and micro level**: “Systems of Innovation” approach is an analytical tool used at several economic levels to understand the relationship between innovation efforts and innovation performance. This approach is already used in other economic perspectives as National [7], Regional [12] and Sectoral innovation systems [13]. The existing relationships between these systems could facilitate or make the establishment of new links between SC players difficult.
- **Demand**: The demand predictability establishes a singular configuration of the SC. The configuration of the SC influences the way a collaborative strategy could be built. With a predictable demand environment the most appropriate SC configuration should be focused on physically efficient processes, whereas with an unpredictable demand a market-responsive-process configuration fits better [14]. Analysing from the innovation point of view, stable demand emphasises incremental innovations instead of radical innovations.

**Internal factors to the innovative SC:**

- **Collaborative culture**: Deals with a collaboration strategy that might be designed, developed, implemented and measured. This is a wide concept that includes other main characteristics, i.e. trust, visibility, sharing processes, etc. Currently, enterprises still organise their internal structures following a functional thinking approach rather than an SC approach [15]. The need to maintain the bargaining power of each department arises from this mentality. Externally the fear of small enterprises at being taken over by larger companies of the SC appears. Although collaboration is one of the objectives of enterprises, there is still a lack of collaborative culture.
• **People:** Even though a collaborative culture is adequate for establishing collaborative objectives, the final responsibility falls on the people that work in this field day-to-day. Social ability, behaviour, motivation, their knowledge and experiences, their role and power in the decision making processes are some of the concepts that should be kept in mind.

• **Powerhouses:** The presence of power centralised in one or two firms in the SC could hinder or facilitate the way to reach agreements. This factor could be omitted if the mutuality principle is applied from the beginning. Collaboration has no purpose if the collaborative strategy partners do not assume mutual risk sharing and mutual benefit from the relationship.

• **Investment capacity:** Economic resources and investment capacity play an important role in the innovation value chain stages. The role is especially important in the second and third stages (idea conversion and idea diffusion) when companies really back some ideas and rule out the rest.

• **ICTs and information exchange:** Successful collaboration requires a change of mind related to information exchange [16]. It is not only the lack of connectivity or the lack of a common platform for information exchange but also the idea that free exchange of data, production schedules and financial information is needed to gain the full benefits of collaboration [17]. According to the degree of development of an innovation and its position following the three stages of innovation value chain, different amounts of information are needed to be shared. Collaborative Planning, Forecasting and Replenishment (CPFR) are the most recent prolific management initiatives that provide supply chain collaboration and visibility. By following CPFR, companies can improve supply chain effectiveness with demand and logistic planning and new product design [18].

• **Physical location:** Despite of the degree of the advanced ICTs used and the volume of information exchanged, the physical location of firms still has influence on this area of analysis. The proximity to the customer or supplier, the personal contact and the possibility of arranging a meeting in the short term could increase the level of trust between firms.

Both external and internal factors affect the SC either directly or indirectly.

### 3.2 STEPS FOR BUILDING COLLABORATIVE STRATEGIES

In the old, linear supply chain, product and information flowed primarily in one direction, and the information itself was limited and not timely. The new internal and external views would provide SC members with a true collaborative innovation capability. The efforts involved in introducing such a collaborative process consist of the 8 steps shown in Figure 2.

![Collaborative innovation process cycle](image)

Figure 2. Collaborative innovation process cycle

(1) Define customers’ value: this is the holistic system view mentioned earlier. This would be accomplished by individual entities to create a total value model. Based on input from the final customers, environmental requirements and laws, overhead policies, the involved parties translate these needs into a value-based measurement system that can be used to evaluate and select concepts that would benefit the entire supply chain.

(2) Members’ innovation focus: this is to identify where and what effort is needed to add the most value.
Establishing a framework for building a SC performance system

(3) Generate ideas: generating a wide variety of concepts is essential by all members in order to add value, i.e. a platform to facilitate communication is then a necessity.

(4) Brief all stakeholders: funneled ideas can be communicated to all stakeholders for feedback.

(5) Evaluate the concepts: Evaluate concepts, using the value-based measures, and select those that can add the most value for final customers.

(6) Assess uncertainties & risks: the main uncertainties and their associated risks are identified and assessed based on the likelihood of their occurrence. Then a plan is developed to reduce the risks as quickly as possible, reducing costs and time for reworking.

(7) Manage collaborative research: primarily oriented towards coordinating research and development activities in order to stimulate and consolidate the supply chain members’ efforts in achieving customers’ needs.

(8) Close the loop: this is to sustain the innovation process and to avoid inertia becoming the norm after a success story or disappointment with a failure preventing the process moving forward again. Here strong leadership is essential.

The above approach should be facilitated by the functional design platform briefly explained above, and also through the development of good practice guidance to help R&D people with different activities required to manage collaborative research projects.

This methodology establishes a common and clear launching point for any collaboration that could be adopted by the SC members. The evaluation system should comply with these basic assumptions in order to maintain the success rate from the premises of inclusiveness, universality, measurability and consistency which will be described in next section.

4. EVALUATION METHODS FOR SUPPLY CHAIN PERFORMANCE

Following Ghalayani and Noble [19] the literature concerning performance measurement has had two main phases. The first phase lasted one century, from the late 1880s until 1980s. In this phase the emphasis was on financial measures such as profit, return on investment and productivity. The second phase started in the late 1980s as result of changes in the world market. Companies began to lose market shares to overseas competitors who were able to provide higher-quality products with lower costs and more variety. New concepts arose from those years: Computer Integrated Manufacturing, Just in Time, Total Quality Management, etc. The strategic priorities of the companies moved from an only low cost vision to quality, product customization…The evaluation methods should also follow the evolution of the “value-added activity” concept. This term arises from the value chain concept in order to describe the value created by one activity and the cost of executing it [20]. Nowadays, the value concept is more extensive, describing the so-called triple P: People, Planet and Profit [21]. These facts lead to change the evaluation perspective from an only financial perspective to a more open evaluation system in which other aspects like social and environmental performance are incorporated.

<table>
<thead>
<tr>
<th></th>
<th>Traditional performance measures</th>
<th>Non-traditional performance measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base:</td>
<td>Outdated accounting system</td>
<td>Company strategy</td>
</tr>
<tr>
<td>Type of indicators:</td>
<td>Mainly financial measures</td>
<td>Mainly non-financial measures</td>
</tr>
<tr>
<td>Intended for:</td>
<td>Middle and high managers</td>
<td>All employees</td>
</tr>
<tr>
<td>Frequency</td>
<td>Weekly or monthly</td>
<td>On-time metrics</td>
</tr>
<tr>
<td>Difficulty:</td>
<td>Confusing and misleading</td>
<td>Simple, accurate and easy to use</td>
</tr>
<tr>
<td>Achievements:</td>
<td>Lead to employee frustration</td>
<td>Lead to employee satisfaction</td>
</tr>
<tr>
<td>Used at operational level:</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Format:</td>
<td>Does not change over time</td>
<td>Changes as change is needed</td>
</tr>
<tr>
<td>Focus:</td>
<td>Monitoring the performance</td>
<td>Improving the performance</td>
</tr>
</tbody>
</table>

Table 1. A comparison between traditional and non-traditional performance measures. [19].

The evaluation systems which are historically considered as effective systems, shared four characteristics [22].

- Inclusiveness: measurement of all relevant aspects. This characteristic demands a holistic vision of the SC. All the relevant aspects of the SC should be considered: strategical, tactical and operational levels.
- Universality: allows for comparison under various operating conditions and benchmarking with other SC.
- Measurability: data required is measurable.
- Consistency: measures must be consistent with organisational goals. It is an alignment between the performance measures and the goals of the SC.

These characteristics are the opposite of the traditional criticism to the SC evaluation systems [23].

- Lack of connection with strategy.
- Focus on cost to the detriment of non-cost indicators.
- Lack of a balanced approach.
- Insufficient focus on customers and competitors.
- Loss of supply chain context, thus encouraging local optimisation.
- Lack of system thinking.

4.1 Evaluation Monitoring System

An evaluation and monitoring system is based on a group of indicators which collect accurate information about the aspects that must be controlled. There are a huge amount of indicators that could be taken into account. This means that the selection of indicators could develop into a complicated task. Key Performance Indicators (KPI) is a technique that facilitates the management and development of many possible indicators. Regarding this issue and from literary reviews it is easy to find different techniques which could be adopted in order to facilitate the management of indicators; [24] support the use of four categories: quality, time, flexibility and costs. Beamon [22] suggested three main concepts: resources, results and flexibility. Vereecke [25] pointed out several categories for measuring performance: costs, quality, flexibility and delivery.

On the one hand, the interests of both academia and practitioners in this field of knowledge are high. On the other hand though, there are not many papers which analyse the role of innovation and collaboration in the social, financial and environmental performances. Furthermore, there is a lack of studies that contribute to the fostering of a holistic vision of the whole SC from the performance analysis [26]. Sheperd and Günter [23] quoted that the work developed by Neely [24] focused on enterprise instead of SC about the analysis of different approaches to performance measurement, between others: the balanced scorecard [27]; the performance measurement matrix [28]; performance measurement questionnaires [29]; criteria for measurement system design [30].

An interesting approach focusing on SC performance was suggested by Simatupang and Sridharan [31] through the Collaborative Performance System (CPS). They referred to CPS as the first step towards developing meaningful SC collaboration. CPS establishes a common language for measuring progress by providing mutual services to final customers and giving the participating members the freedom to improve rapidly or even immediately the SC operations as a whole. This approach maintains a traditional view of SC as an operational function. Therefore, innovation is not the driver. When a SC is able to deliver the best value to their final customers, it will generate strong demand and build customer loyalty to its products and services.

Collaborative innovation among SC members becomes an essential approach in order to create sustainable competitive advantages. However, the strategies and approaches in innovation adopted in the 80s and 90s, are no longer sufficient. Supply chain members must involve in effective experiments and innovation to reinvent the way they create value, since conventional business does not produce the desired results in fast clockspeed chains [32].

4.2 Performance Evaluation System

The management of the enterprise sustainability used to be based on the maximisation of the benefits thanks to the increase of sales or/and the reduction of manufacturing and distribution cost. Historically companies do not consider the effects of their production neither at social nor at environmental levels. Nowadays this behaviour has no sense so a change of mind is needed. Although the measurement of social issues could be a difficult task, the companies of the same SC should analyse the chain as a place of interaction. In this new space of interaction people could generate new ideas, information flows without obstacles and so on. There are two main issues to develop: fostering creativity and strengthening trust through the whole SC. SC becomes in an intangible good rather than a mere infrastructure.
Establishing a framework for building a SC performance system

Regarding the environmental issues, the best way of establishing the correct indicators consists in analysing how the environmental management has changed over the past. The environmental indicators usually do not reflect the damages generated by the companies on the environment [33]. However, companies have to cover the environmental area due to the pressures by society.

5. CONCLUSIONS

In the dynamic environment of most supply chains, collaborative innovation is a fundamental strategy for sustainable and profitable growth. Growth through collaborative innovation process is more efficient than setting up an independent innovation. Managing supply chain by cost can lead to a downward spiral and has limited potential for success. Instead, creating an innovative supply chain with a collaborative innovation culture can offer unlimited potential for success. Managing only the supply chain cost limits the freedom to innovate; whilst leading towards collaborative innovation opens minds. This is a positive experience for everyone – however it involves its own complexities to manage. If the collaborative innovation strategy is built from the suggested framework, the results could cover a wide range of results.

SC performance evaluation systems have traditionally been focused on financial issues, although innovation characteristics demand a change of mind. Following this idea there are two main points that should be addressed: Firstly, it is necessary to develop measures of SC relationships and the SC as a whole and secondly, the number of qualitative and non-financial measures should be increasingly included in the evaluation systems. The motivation for measuring other areas apart from the economic field should come from the satisfaction of stakeholders. Taking care of the environmental and social affairs is to the advantage of the SC through the better understanding of the customer needs, motivated employees, the SC image, etc.

REFERENCES

Optimization of Multi-Echelon Supply Chain Networks Using Quantum Inspired Particle Swarm Methodology

M Bachlaus¹, F F Chen²*, W.H.A. Wang³ and D.W Kim⁴

¹,² Department of Mechanical Engineering
University of Texas at San Antonio
San Antonio, TX, 78256, USA

³ Department Industrial Engineering
& Enterprise Information, Tunghai University, Taiwan

⁴ Department of Industrial and Information Systems Engineering
Chonbuk National University, Korea

ABSTRACT

This paper presents a novel approach that integrates the intangible and tangible factors to model the resource assignment problem in a product driven supply chain. The problem has been mathematically modeled as a multi-objective optimization problem with the objectives of maximum profit, best quality, early delivery, and volume flexibility. In this research, product characteristics have been associated with the design requirements of a supply chain. Different types of resources have been considered each differing in its characteristics, thereby providing various alternatives during the design process. The aim is to design integrated supply chains that maximize the weighted sum of the objectives, while weights being decided by the desired product characteristics. The problem has been solved through a proposed Quantum inspired Particle Swarm Optimization (QPSO) metaheuristic. It amalgamates particle swarm optimization with quantum mechanics to enhance the search potential and make it suitable for integer valued optimization. The performance of the proposed solution methodology and its three variants has been authenticated over a set of test instances. The results of the study and the insights derived through it validate the efficiency of the proposed model as well as the solution methodology in the underlying problem.

1. INTRODUCTION

A supply chain, in general, comprises a number of business entities including suppliers, manufacturers, distributors, vendors and their respective haulers. Therefore it represents a distributed system which requires joint strategic planning to ensure effective performance (Poirier and Quinn, 2004). The decisions that take place in a supply chain have been classified into four levels of hierarchy - strategic, tactical, operational, and real-time levels. The most influential among these levels is the strategic level, which deals with the resource assignment, flow of goods, cost-effective location of facilities, etc. The success of the decisions taken at other levels depends to a large extent, upon the decisions that have taken place at this level, thus it can be considered as most important level. However, the decision making at this level is a difficult process as it involves the resolution of a complex resource allocation problem which involves the selection of different business entities (out of the available ones) based on various tangible and intangible factors.

This paper focuses on the strategic decision level of the supply chain and thereby considers the resource assignment problem in a product driven supply chain. The operational effectiveness of such a supply chain is mainly determined by the involvement of tangible (equipments, plants, fleets, hardware) and intangible factors (organizational processes, skills, know-how, quality, reputation). Since tangible factors are measured in quantitative terms, these can easily be incorporated into the design consideration of the resource assignment problem. However, many intangible factors cannot be measured directly. One example of this is quality. A number of methodologies (Saaty, 1994; Meade and Sarkis, 1998; Wang et al., 2004) are suggested in the literature to quantify the intangible factors. Most of these methodologies assign relative scores to each of the available alternatives and rank them on the basis of these scores. However, an effective integration of intangible factors with the tangible ones is still a challenge faced by the research community. In this paper, a multi-objective optimization model has been formulated in order to integrate these two factors into the resource assignment problem of a product driven supply chain. The traditional Analytical Hierarchy Process (AHP) employs crisp values to score the criteria, which is not an exact and realistic method to specify the decision maker’s feelings. Thus, this paper introduces the application of mathematical tools of fuzzy set theory (Zadeh, 1974) into the decision making model. The proposed approach facilitates the notion of expressing the...
decision maker’s opinion in linguistic variables such as high, medium or low which are thereafter translated into the mathematical form by presenting them in a form of triangular trapezoidal fuzzy numbers.

In order to solve the aforementioned optimization problem, this paper proposes a novel metaheuristic for constrained or unconstrained optimization problems entitled Quantum inspired Particle Swarm Optimization (QPSO). The skeleton of the proposed QPSO metaheuristic encompasses an encoding schema which amalgamates the quantum philosophy (Rieffel and Polak, 2000) with the particle swarm strategy (Kennedy and Eberhart, 2001). The encoding schema offers advantages such as thorough representation of the search space, reduced dependence on population size, and enhanced search capability (Grover, 1996), while the characteristics of particle swarm aid in efficiently orienting the search towards promising regions of the search space. The metaheuristic and its three variants (QPSO, QPSO-with random inertia weights, QPSO-time varying inertia weights and QPSO-time varying acceleration coefficients) have been applied to solve the aforementioned supply chain resource allocation problem and the results were thoroughly analyzed to draw useful insights.

2. LITERATURE SURVEY

A comprehensive study of the literature reveals that there has been much focused work in the area of supply chain design and modeling. Arntzen et al. (1995) design a production, distribution, and vendor network for a global supply chain model. Their model addresses the issues related to reengineering an existing supply chain, however it does not consider the efficiency of supply chain network processes. Erenguc et al. (1999) proposed a model of a supply chain network which is composed of three stages: supplier network, producer network and distributor network, and defined the nature of the relationship between each stage. An integrated multi-objective supply chain model for strategic and operational supply chain planning has been developed by Sabri and Beamon (2000). Yan et al. (2003) proposed a strategic production-distribution model for supply chain design with consideration of bills of materials for operational processes. For an extensive review of strategic production-distribution models in a global supply chain environment, readers are referred to Govil and Proth (2002).

While scanning the literature, it can be easily concluded that there is not a holistic method to measure the operational quality of resources in a product driven supply chain. Although few attempts have been made in the literature to incorporate the intangible factors into the design consideration, researchers are still investigating more-efficient approaches that can cope with these drawbacks. Other important observations obtained from the existing body of literature are:

1. Intangible factors like the quality of a product can be an effective measure to determine the operational performance of a product driven supply chain. However, till now these factors are not integrated in the design consideration of resource assignment problems at the strategic decision making level.

2. Since, a supply chain design problem is NP-hard (Amiri, 2006), there is a need for the development of an efficient solution methodology that can identify the optimal or near-optimal solution in minimum computational time.

3. PROPOSED MATHEMATICAL MODEL FORMULATION

In this paper, a product driven supply chain problem comprised of four different echelons i.e. suppliers, haulers, manufacturers and vendors has been considered (Figure 1). All of these echelons are linked together with a forward flow of material and backward flow of information. At the strategic level, this paper aims at reducing early delivery and cost while, improving product quality and volume flexibility. The typical supply chain network structure considered in this paper is aimed at the integrated design of supply chains for manufacturing and delivery of different grades of a product. It has been proposed that the selection of an appropriate type of supply chain should be driven by the characteristics of the product(s) an organization is manufacturing. The supply chain model formulated in this study constitutes different kinds of suppliers, manufacturers, their respective haulers, and different vendor locations at which the final product is to be delivered. The scenario considered here is based on the design of a one-to-one business process chain where each vendor desires a specific grade of product which is produced and delivered from a selected manufacturer, each manufacturer only orders parts from one supplier, and each supplier and manufacturer can select a hauler available to them to transport their products.

The goal of optimizing resource assignment is to identify the optimal number of upstream suppliers, manufacturers, downstream vendors and their corresponding haulers. Thus, a supply chain ensuring certain profit,
early delivery, better quality, and volume flexibility can be designed. In particular, supply chains with varying weights of the objectives have to be designed with appropriate allocation of resources.

Owing to the inefficient and unreliable weighted sum method as suggested in the literature, an approach that generalizes the aggregating schemes into some “ideal” using mathematical tools of fuzzy set theory has been proposed. This approach utilizes the fundamentals of fuzzy set theory and it can be implemented in a large number of applications. A brief introduction of the suggested approach is discussed in the following sub-section.

3.1.1 Steps of fuzzy based multi-criteria decision making model
There are seven steps in the proposed approach:
Step 1: Form a weighting team of decision makers to evaluate the performance metrics associated with the given problem.
Step 2: Determine the evaluation criteria of suppliers, manufacturers and haulers. These criteria are adopted from SCOR.
Step 3: Set proper linguistic scales (e.g. very high, high, medium, low, and very low) and ask decision makers to give their opinions on the relative importance of the evaluation criteria by pair-wise comparison.
Step 4: Convert the linguistic variables into triangle fuzzy numbers to get every member’s fuzzy reciprocal matrix (Table 1).
Step 5: Aggregate the decision maker’s fuzzy reciprocal matrices by geometric means and form the final aggregated fuzzy reciprocal matrix.
Step 6: Take the geometric row means of every performance criteria and normalize it to obtain its local weight.
Step 7: Translate each final triangle fuzzy score of performance criterion into a crisp value for ranking purposes. There are several de-fuzzification methods in the literature. This paper, however, employs Opricovic and Tzeng’s (2003) formula because of its effectiveness in solving combinatorial optimization problems. Based on these steps, the quality associated with each supplier, manufacturer and hauler can be determined once the assignment of resources has been completed. The aggregate quality of the supply chain can then be calculated by taking the average of the score of all the members involved in the supply chain.

![Figure 1: The network of product driven supply chain](image)
Based on the network graph shown in Figure 1, the multi-objective optimization model for resource assignment problem in product driven supply chain can be written as:

$$\text{Max Obj} = \sum_{v=1}^{V} \left( w_{1v} \cdot \text{norm}(P_v) + w_{2v} \cdot \text{norm}(AT_v) + w_{3v} \cdot \text{norm}(Q_v) + w_{4v} \cdot \text{norm}(VF_v) \right)$$  \hspace{1cm} (1)

### 3.2 Proposed Multi-objective optimization model

The individual objectives (i.e. profit, early delivery, quality, and volume flexibility) are mathematically modeled here. Operational quality assignment to resources is done by the approach outlined in Section 3.1.1. The details of the remaining objective formulation are given below:

#### 3.2.1 Profit

The profit of product driven supply chains considered here is given by the difference of the price of the product and the total cost incurred during the manufacturing and delivery of the product (equation 2). The total cost of the four echelons of supply chains is calculated at each stage and can be mathematically represented.

$$P_v = UP_v \cdot \text{Vol}_v - C_v$$  \hspace{1cm} (2)

#### 3.2.2 Early delivery

Early delivery refers to the time left ahead of the scheduled delivery of the product. It can be mathematically modeled as the difference between the observation period of all vendors and the maximum delivery time available. Equation (3) denotes the ahead-time of delivery calculation whereas, the maximum delivery time can be determined

$$AT_v = OT_v - DT_v \hspace{1cm} \forall \ v$$  \hspace{1cm} (3)

#### 3.2.3 Volume Flexibility

Volume flexibility can be defined as the minimum of the difference between the suppliers, haulers, and manufacturer’s capacity with the total volume of products ordered by vendor $v$ (equation 4). First, the individual flexibility of the suppliers, haulers and manufacturers will be calculated. Thereafter, the minimum of the obtained flexibility would be considered as the final volume flexibility of the product driven supply chain. Therefore, volume flexibility can be written as

$$VF_v = min\left( SF_v, SHF_v, MF_v, MHF_v \right)$$  \hspace{1cm} (4)

#### 3.2.4 Constraints

Since the proposed mathematical model formulates a one-to-one product driven supply chain, the following constraints (Equations 5, 6, 7 and 8) are modeled in order to check that only one supplier, manufacturer, and their respective single hauler will be selected from the available resources. Equations 9 and 10 ensure that the continuity will remain in the supply chain at each echelons. If a particular type of supplier is not selected in the process chain, it indicates that in the next echelon that supplier would not be taken into account.

$$\sum_{i=1}^{S} a_{iv} = 1 \hspace{1cm} \forall \ v, i$$  \hspace{1cm} (5)

$$\sum_{i=1}^{S} y_{ivm} = 1 \hspace{1cm} \forall \ v, i, m$$  \hspace{1cm} (6)

$$\sum_{m=1}^{M} b_{mv} = 1 \hspace{1cm} \forall \ v$$  \hspace{1cm} (7)

$$\sum_{j=1}^{MH} l_{mvj} = 1 \hspace{1cm} \forall \ v, m$$  \hspace{1cm} (8)
Equation 11 depicts that the cost of the products ordered by vendor \( v \) should be less than or equal to the total cost of the products manufactured by the resultant supply chain of vendor \( v \). Equation 12 restricts the maximum delivery time of the products ordered by vendor \( v \) to be less than or equal to their observation period. It is also imperative to ensure that the maximum capacity of suppliers, their respective haulers, and manufacturers should be more than the volume of the products ordered by vendor \( v \). This constraint at each echelon is modeled in equations 13, 14, 15 and 16, respectively. In the next section, the proposed solution methodology is detailed along with the explanation of the implementation procedure to the underlying problem.

4. THE PROPOSED QUANTUM INSPIRED PARTICLE SWARM OPTIMIZATION METAHEURISTIC

The proposed metaheuristic derives its governing traits from two ideologies: conduct of particles in the swarm and quantum mechanics. While the socio-cognitive behavior of particles acts as the state transition mechanism and enables the search to move towards optimal solution, quantum inspired characteristics provide stochastic characteristics and ensure the convergence towards global optimum in the long run. The proposed QPSO is simultaneously characterized by rapid convergence and a global search capability.

The proposed approach differs from the traditional particle swarm algorithm in its representation of a particle’s position. In this study, it is considered as stochastic instead of being deterministic. A fitness based probability update mechanism is also devised and incorporated to enhance the probability of selection of those members whose selection have shown better performance.

The acquisition of principles from quantum physics in a traditional particle swarm approach offers distinguished advantages such as

1. One individual of QPSO can represent many states at the same time. This property reduces the dependence of the search on the population size, since a small population can represent the entire search space effectively while maintaining the required diversity and transition pressure. However, this reduces the computational burden.

2. QPSO can be utilized for optimizing complex combinatorial optimization and integer programming problems.

It ensures positive kernel variation and through the state transition mechanism of PSO (cognitive and social search components), it leads the search towards global optimality.

A. Encoding Schema: The proposed encoding schema, to a certain extent, is based on the concept of qubits and superposition of states of quantum mechanics. In the proposed approach, real numbers, representing probabilities of the selection of an entity, are used for encoding. The technique is illustrated through an example shown in Figure 2. Consider a basic problem of selecting 2 suppliers out of the 6 available ones for two different supply chains. The
selection of two suppliers is carried out in two phases. The stochastic encoding scheme proposed for this is shown in Figure 2(a). The six bits in the string represent different suppliers with the values in those bits representing the probability scores of selection. In the first phase, these values are first normalized to sum to one, as shown in Figure 2(b). The values obtained represent the normalized probability of selection of each supplier and the first supplier could be selected from it using the roulette wheel selection strategy. Suppose that supplier 2 is selected as the first supplier. Now, at this stage, second supplier needs to be chosen from remaining 5 suppliers (Figure 2(c)). This constitutes the second phase in which the individual probability of selection of the remaining suppliers (Figure 2(d)) is again calculated and thereafter the roulette wheel selection is applied. Thus, the number of phases depends upon the total number of supply chains required to be designed. Evaluation of the objective function can be done once the process of string decoding has been completed. This methodology, apart from providing other benefits, ensures that the constraints of the problem are not violated. The proposed methodology is generic and with suitable modifications can be applied over a wide range of problems.

B. Fitness based Update rule: This mechanism has been devised to provide proportional benefits to those selections or allocation which produce better results. After evaluation, the fitness value of the decoded string is compared with the fitness value stored in the particle’s memory. If the current fitness value is found to be better than the value stored in memory, the probability values in the bits corresponding to the selected ones is proportionately increased using the equations,

\[
Incr_y = \frac{Fit_i}{PN \times C}
\]  

\[
x_y = x_y + Incr_y
\]

where, \(x_y\) represents those bits which have been selected, \(Incr_y\) denotes the amount of increment in those bits which are selected, \(Fit_i\) is the fitness value of the particle \(i\) after decoding, \(PN\) is the phase number, and \(C\) is a constant which is used here as a scaling factor.

In this research, the proposed approach has been implemented on the resource allocation problem with three proposed variants, namely, QPSO with random inertial weights, QPSO with time varying inertial weights and QPSO with time varying acceleration coefficients. The concept of variants is derived from Ratnaweera et al. (2004) since the modifications proposed by them in PSO have established better performance.
5. RESULTS AND DISCUSSIONS

In this section, results of the proposed methodology and its variants on the resource allocation problem have been detailed. Eight instances that correspond to the problem faced by a leading steel manufacturing group have been simulated. The problem encompasses the design of supply chains within the available resources for the manufacture of several similar kinds of product. The problem design has been detailed in Table 1. All the problem instances consist of three types of resources (available suppliers, manufacturers and their respective haulers), which are classified as A, B and C, respectively. Resources of Type A have a high level of quality, and therefore a corresponding high manufacturing cost, and a slower manufacturing time. Type B resources are having these values comparatively inferior. Finally, the least values are assigned to the resources of Type C. The product driven supply chains to be designed are divided into three categories: quality driven (maximization of quality), cost driven (minimization of cost or maximization of profit) and homogenous combination of the other two. The weights given to the objectives of quality, cost, flexibility and early delivery are taken as 0.4, 0.2, 0.2; 0.1, 0.4, 0.25, 0.25; and 0.25, 0.25, 0.25, 0.25, respectively for the aforementioned three kinds of supply chains examined in this study.

5.1 Experimental Results

The proposed quantum inspired particle swarm optimization algorithm and its three variants namely, QPSO-RandIW, QPSO-TVIW, and QPSO-TVAC have been implemented over all the problem instances. The parameter setting used for all the algorithms have been listed in Table 3 and the obtained results are provided in Table 4. As is evident from Table 3, all the algorithms utilized similar parameter settings to ensure that the comparison is being carried out on same platform. The fitness function is taken as the weighted sum of normalized objective function values, the weight being assigned as per number of supply chains to be designed. The number of fitness function evaluations has been taken as the termination criteria. A hundred independent runs were performed for each problem instance for all the four algorithms concerned in this paper. Table 2 lists the mean and standard deviation values of the fitness of the best particle found.

Table 1: Details of problem instances undertaken in the case study

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Type of suppliers ($S_i$)</th>
<th>Number of suppliers of each type ($S$)</th>
<th>Number of haulers for each suppliers ($SHM_{si}$)</th>
<th>Number of manufacturers ($M$)</th>
<th>Number of haulers for each manufacturers ($MH_{im}$)</th>
<th>Number of supply chains to design</th>
<th>Type of supply chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>Q</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>Q, C</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>Q, C, H</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>Q, C, H</td>
</tr>
</tbody>
</table>

* Q = Quality Driven, C = Cost Driven, H = Homogenous

It is evident from Table 2 that results obtained by all four algorithms are comparable; however, QPSO-TVAC shows the best performance with the highest values of average fitness and the minimum standard deviation than the rest of the algorithms for the majority of the problem instances considered. The low values of standard deviation obtained by the QPSO-TVAC demonstrate its high repeatability and good convergence characteristics. The second best results were shown by QPSO-TVIW, followed by QPSO and the worst performance registered was that of QPSO-RandIW. The convergence trend obtained with the four algorithms on problem instance 1 is shown in Figure 3. The percentage contributions of different types of resources (A, B and C) in the optimized design of different grades of supply chain for the test instances considered in this paper are reported in Table 3. It is evident from the table that in the design of the quality driven supply chain and the cost driven supply chain, the major contributions are from the resources of type A and type C respectively. The results indicate that the quality objective in a supply chain can only be fulfilled when most of the partners are quality conscious. However, restrictions enforced by cost criteria limit their proportion in the optimum design. The results also reveal that in the homogenous design of supply chains, resources of type B...
occupy a major proportion, which is contrary to the general assumption of equal proportions of all the three types of resources in its optimum design.

Table 2. Results for all algorithms on benchmark problems (mean of 100 runs and standard deviations (Std. dev)). For each problem, the best performing algorithm(s) is emphasized.

<table>
<thead>
<tr>
<th>Problem No.</th>
<th>QPSO</th>
<th>QPSO-RandIW</th>
<th>QPSO-TVIW</th>
<th>QPSO-TVAC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
<td>Std. dev.</td>
</tr>
<tr>
<td>1</td>
<td>0.9696</td>
<td>0.001285</td>
<td>0.9690</td>
<td>0.001293</td>
</tr>
<tr>
<td></td>
<td>0.9712</td>
<td>0.001280</td>
<td>0.9716</td>
<td>0.001279</td>
</tr>
<tr>
<td>2</td>
<td>0.9461</td>
<td>0.001352</td>
<td>0.9456</td>
<td>0.001368</td>
</tr>
<tr>
<td></td>
<td>0.9479</td>
<td>0.001358</td>
<td>0.9473</td>
<td>0.001361</td>
</tr>
<tr>
<td>3</td>
<td>1.5161</td>
<td>0.001794</td>
<td>1.5142</td>
<td>0.001802</td>
</tr>
<tr>
<td></td>
<td>1.5191</td>
<td>0.001765</td>
<td>1.5224</td>
<td>0.001782</td>
</tr>
<tr>
<td>4</td>
<td>2.0875</td>
<td>0.002123</td>
<td>2.0901</td>
<td>0.002067</td>
</tr>
<tr>
<td></td>
<td>2.0965</td>
<td>0.002064</td>
<td>2.0992</td>
<td>0.002041</td>
</tr>
<tr>
<td>5</td>
<td>2.0962</td>
<td>0.002186</td>
<td>2.0975</td>
<td>0.002193</td>
</tr>
<tr>
<td></td>
<td>2.1006</td>
<td>0.002164</td>
<td>2.1073</td>
<td>0.002169</td>
</tr>
</tbody>
</table>

Table 3. Percentage contribution of different types of members in the optimum supply chain configuration of different problem instances

<table>
<thead>
<tr>
<th>Problem instance</th>
<th>Type of design</th>
<th>% contribution of type A</th>
<th>% contribution of type B</th>
<th>% contribution of type C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem instance 1</td>
<td>Q</td>
<td>0.79</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Problem instance 2</td>
<td>C</td>
<td>0.184</td>
<td>0.298</td>
<td>0.518</td>
</tr>
<tr>
<td>Problem instance 3</td>
<td>Q</td>
<td>0.82</td>
<td>0.153</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.267</td>
<td>0.245</td>
<td>0.488</td>
</tr>
<tr>
<td>Problem instance 4</td>
<td>Q</td>
<td>0.74</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.123</td>
<td>0.317</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>0.258</td>
<td>0.456</td>
<td>0.286</td>
</tr>
<tr>
<td>Problem instance 5</td>
<td>Q</td>
<td>0.71</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.292</td>
<td>0.333</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>0.25</td>
<td>0.42</td>
<td>0.33</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS AND FUTURE WORK

In this paper, a model of resource allocation for a product driven supply chain has been considered. The supply chain is modeled as a network where, nodes represent resources in the network. The objective of the study was to design a supply chain with specific requirements and to maximize the integrated objective comprising of quality, profit, volume flexibility and early delivery. In order to solve the problem, a novel strategy based on particle swarm optimization and quantum mechanics is devised. This research raises several questions, of which, the most important and those requiring further research are listed. First, the model has been formulated for a one-to-one supply chain and need to be enhanced to include the possibility of multiple resource selection at each stage. For instance, a vendor may desire to procure a raw material from two different suppliers in order to have more flexibility and minimize disruptions. Second, inclusion of other intangible objectives such as customer service satisfaction level, skills etc. is worth to be explored. The inclusion of web based enterprise system and its effective integration with supply chain design is also a topic for future research.
Optimization of Multi-Echelon Supply Chain Networks Using Quantum Inspired Particle Swarm Methodology

![Convergence trend of the proposed algorithms](image)

**Figure 3:** Convergence trend of the proposed algorithms

**REFERENCES**

Exceptions management in the supply chain

Esther Álvarez*, Fernando Díaz†, Miguel A. Larrinaga‡

1 Faculty of Engineering
University of Deusto
Bilbao 48012, Spain

2 Faculty of Economics and Business Administration
University of Deusto
Bilbao 48012, Spain

A supply chain is composed of a set of different agents such as suppliers, manufacturers and distributors. In order to manage a supply chain in a successful way, it is necessary to reduce some key aspects like lead times and stock levels, since they contribute to increase the global cost of the supply chain. This research aims at building a proactive exceptions management solution that operates in real time by monitoring possible disturbances and processing them on a wide basis, thus considering their impact on the overall supply chain. In this way, a suitable synchronization of the different production schedules will be achieved, thus leading to a global competitive advantage over competitors. The selected approach will make use of distributed control, where decisions are made at each node taking into account the information interchanged with other nodes of the supply chain. This means that each plant will be autonomous as regards production planning and scheduling but at the same time will take advantage of the common information of the supply chain, so as to allow a dynamic collaboration between the agents. Some expected benefits of the system are the following: a better information visibility, a fast notification of disturbances throughout the supply chain, a synchronized production and a collaborative production scheduling. This research is being supported by Grant PI2008-08 from the Bask Government in Spain.

1. Introduction

In today’s highly competitive global markets, within a global economic downturn, it is vital not only to attract customers but also to retain them. When a customer buys a product or service and he or she receives it, this consumer only sees the last link in a chain of agents that guarantees the quality of the final product or service. The overall management and coordination of all these agents is revealed as a cornerstone for sustainable growth of the company. We are moving from managing a company to managing a supply chain. Companies don’t compete; supply chains compete [1].

In order to do this effectively it is necessary to try to reduce as much as possible some aspects such as delivery times and inventory levels, as these are translated into costs for the company. A smooth communication between players who are part of it is essential. This makes it possible to synchronize the different production schedules, gaining a valuable competitive advantage. This synchronization is complicated because it is necessary to break cultural, personal, organizational, information and technology barriers.

One of the solutions to improve the supply chain is to identify the information that should be exchanged between nodes and develop methodologies for the collaborative production planning. The main objective is not only to manage the value chain of each individual company, but also to manage the overall value chain in the extended company, since it is necessary to analyze the implications of the changes occurring at a point affecting other points in the supply chain. The intention is to connect the production planning of the different agents in the supply chain to optimize the production planning of the whole supply chain.

Some of the most significant advantages of the system that are expected to be achieved are to improve the visibility of information throughout the supply chain, to early notify disturbances throughout the supply chain and to contribute to synchronized production and collaborative scheduling.

The objectives of this research project are the following:

* Corresponding author: Tel.: (+34) 944139000; Fax: (+34) 944139101; E-mail: ealvarez@eside.deusto.es
• To identify the information to be exchanged between the different agents in the supply chain.

• To collaborate with suppliers in order to ensure the availability of raw materials when new unplanned materials are required.

• Typical exceptions must be identified in order to provide a solution automatically and in real time. Changes can come from three sources which define three types of exceptions: [2]
  o Inside the company itself (internal exceptions).
  o Exceptions coming from suppliers (supply side).
  o Exceptions coming from customers (demand side).

• To early warn customers about any changes that could affect order fulfillment.

• To increase the flexibility of the solution by using resources at different plants. When there is a network of plants that can provide alternative resources, it may happen that a particular product can be manufactured in different plants. Once a job is allocated to a plant, it may happen that an unexpected event that delayed the work justifies its reassignment to another plant in real time, or outsourcing it dynamically [3].

This paper is organized as follows. First, a review of related research is provided. Then the architecture of the system is shown and explained. Finally, a few concluding remarks and possible directions of future research are discussed.

2. STATE OF THE ART REVIEW

A state of art survey on all the aspects related to the objectives of the project is presented in the following sub sections. A classification has been made in two groups: research studies in collaborative supply chain management and research studies in production scheduling.

2.1. RESEARCH STUDIES IN COLLABORATIVE SUPPLY CHAIN MANAGEMENT

The goal of a supply chain is the maximization of the overall value generated. In order to achieve this goal a fluent flow of information through the supply chain is necessary. However, an unplanned demand oscillation will cause distortions on the supply chain. This distortion is known as the “bullwhip effect” [4]. This effect was considered normal in the past but the negative impact that it causes in business, (e.g. excess of inventory, shipping cost, quality problems, etc.) is the reason for present research in this area.

Since research started to grow, there have appeared different models in order to represent the way a supply chain has to be managed. Of all these models there are two that are the most used nowadays.

On the one hand, there is The Supply Chain Operations Reference Model, commonly known as SCOR [5]. It is a diagnostic tool for the Supply Chain Management that spans all customer interactions, including order entry to payment, all the products transactions, all around the supply chain and all market interactions. Some authors showed [6], through a simulation study, that to enjoy the full benefits of collaboration, practitioners should focus more on synchronization than just on information visibility.

On the other hand, the CPFR model [7] offers a general framework by which a buyer and a seller can use collaborative planning, forecasting and replenishing processes in order to meet customer demand. Buyers and sellers are involved in four collaboration activities, namely Strategy and planning, Demand and Supply Management, Execution and Analysis.

These models have been used by different institutions as a reference in order to create different models to represent a supply chain.

The CO-OPERATE project [8] aims at improving the overall goal of the supply chain by creating a communication infrastructure between companies. This infrastructure enables a collaborative production planning, multi sourcing coordination, process visibility and exception handling reducing the “bullwhip” effect thanks to information sharing.

According to [9], the relationship between utilization, optimum time and work in progress plays a very important role in the design of a supply chain and a manufacturing system. This research shows that inventory or work load variability cause capacity waste, low machine utilization rates and long optimum time.
Some authors have explored the possibility of integrating different stages of the value chain that have traditionally been solved separately. The approach in [10] investigates production planning in a multi-plant company where each plant can serve different customers, so when a new order arrives, a decision must be made as regards plant allocation. The authors develop the model based on a linear programming algorithm in order to obtain the aggregated plan and the detailed plan simultaneously. Thanks to it, it is possible to beat the hierarchical methods’ limitations that usually cannot find a possible solution or a sufficiently optimized one. Another approach [11] states that companies have tried to optimize production and distribution systems separately. Therefore, it presents a solution for the integration of production planning and distribution, to improve their productivity by incorporating a multi-stage, multi-use, multi-item and multi-period plant. The main goal is to maximize the total net profit. The study shows that the integrated approach substantially outperforms the approach that considers both stages of planning and distribution separately. In order to do this, the author uses a heuristic in the solution of production planning and distribution in an integrated manner. However, this model is not computationally feasible when the number of plants, customers, consumer product and periods is high.

Other research studies have taken advantage of creating schedules with auctions that fit naturally into a distributed architecture. In [12] an auction-based solution is presented where jobs are planned in a distributed way. The resulting problem contains a variable number of operations, different times of process etc. Each task has to be done following a set of operations where each one requires a specific machine with contiguous time slots. This type of solution is not often used because it is impossible to obtain an exact solution and the generated planning is not real. In [13] an algorithm based on bids for the sequencing of real-time flexible manufacturing systems with alternative routes is presented. The effectiveness of the proposed framework is verified by comparing it with a number of priority rules using simulation. The analysis of the results showed that the approach based on auction outperformed the rules of priority in almost all measures of performance. Nevertheless, the auction-based method is not well accepted in industry because of the unpredictability of the system performance, poor quality of schedules generated, and no theoretical basis for computing the bids and bid evaluation.

In [14] an interesting model prepared to manage Complex Adaptive Systems is presented. This type of model reacts before unexpected events that might appear by continuously revising both planning and scheduling. Other authors [15] propose a new algorithm (NEG) based on negotiation in order to solve the distributed projects problem. This algorithm allows all nodes in the supply chain to make decisions independently but it also makes the most of shared information to improve the quality and effectiveness of the supply chain. A study shows that NEG works better than a heuristic centralized algorithm and can obtain solutions in a short period of time.

Therefore, two successful contributions to SCM are the SCOR and CPFR models that provide tools to reduce the bullwhip effect and best meet customer demand. Besides, the CO-OPERATE project enables collaborative production planning and exception handling by means of a creating a common information infrastructure in the supply chain. Despite all contributions in this research area, there is a lack of studies that focus on synchronising local planning and scheduling solutions in real time to attain a global optimal solution. Improved tools are required to handle exceptions that optimise globally rather than locally over the extended enterprise.

### 2.2. Research Studies in Production Scheduling

Job shop scheduling is a combinatorial optimization problem of class NP-complete [16]. In practice, the exact resolution of this type of problems is impossible in reasonable execution times, so approximate methods must be used.

First, we can use heuristic methods specifically designed for the resolution of the considered problem. In case of the job shop scheduling problem, the most common heuristic method is that of the classic dispatching rules (DR), especially in a dynamic industrial context, where work orders arrive at the shop floor randomly in time. Due to their easy implementation and considerably small computational needs it turns out to be a very popular technique [17]. The method consists of sequencing work orders according to a priority order established by the dispatching rule or rules used but in general several executions of the procedure are needed in order to reach valid results. Much research has focused on selecting the dispatching rule that best optimizes a wide range of performance measures related to work orders (such as due dates and tardiness) and related to the shop floor (such as performance and utilization). This problem of selecting the best DR for a given performance measure continues being a very active research area [18], [19], [20], [21], [22].

Another class of methods is based on the construction of general algorithms of combinatorial optimization on several possibilities:
Path methods. It is a class of approximate optimization algorithms starting with a random initial complete solution, which is progressively transformed into new solutions by means of partial modifications until the best possible solution is reached in a feasible time. This procedure can be interpreted as a process of search across a path in the solution space. Among others, the most important ones are local search [23], tabu search [24], and simulated annealing [25], as well as the hybrid method of simulated annealing-tabu.

Population methods. It is a class of approximate optimization algorithms similar to the path methods in the sense that they operate on complete solutions that are being transformed, but instead of considering only one solution in each iteration of the algorithm, a set of solutions or population is used. For that reason, they provide a multiple parallel search process in the solution space of the problem. Among others, the most important ones are genetic and evolving algorithms [26], the hybrid ones coming from them, named memetic algorithms [27] and genetic algorithms with simulated annealing, and independently the particle swarm optimization [28].

Constructive methods. They are approximate optimization procedures that construct a solution of the problem assigning values to the variables in a sequential way. In this type of algorithms we can find the greedy randomized adaptive search procedure (GRASP) [29], and the ant colony optimization [30].

Among all these general algorithmic techniques of combinatorial optimization, one of the most frequently used for the job shop scheduling problem in the last years has been specifically that of genetic and evolving algorithms. The following references are only a sample of the considerable work realized in this field.

Some researchers have centered on combining genetic algorithms with methods of problem decomposition (divide and conquer technique). For example, in [31] a resolution method for the job shop scheduling problem based on the combination of a standard genetic algorithm with the concept of group technology, GT is presented. The main idea of the method is to reduce the complexity of the problem by separating work orders in groups in accordance with the similarities among the corresponding parts, and to make the program of every group independently. Likewise, some researchers [3] develop a model of assignment and programming to study the impact that flexibility in manufacturing produces in such important aspects of production as due dates fulfilment and machine utilization. In particular, a resolution method of divide and conquer type is contributed, in which the job shop scheduling problem is decomposed in two sub problems: the assignment of machine to operations (grouped in types) and the assignment of work orders to machines. A later paper of the same authors [32] presents a similar approach, introducing as innovation resource constraints to limit machine flexibility.

Other researchers have focused on stochastic job shop scheduling, a special instance of the problem, and have applied genetic algorithms to solve it. For example, a study proposes a resolution method for the stochastic job shop scheduling problem based on a genetic algorithm combined with stochastic mathematical programming [33], in which the fitness function (objective) is considered to be a fluctuation that depends on the stochastic conditions specified by the probability distributions of the random variables of the problem. Likewise, other researchers propose a resolution method for the stochastic job shop scheduling problem based on a modified genetic algorithm with search area adaptation, mGSA, and independent from the crossover and mutation parameters [34].

Genetic algorithms have been also utilized in the resolution of the open job shop scheduling problem, a special case of production scheduling in which every work order consists of m manufacturing operations that can be processed in any order. Some authors develop a heuristic genetic algorithm to solve this scheduling problem [35]. This genetic algorithm always obtains feasible solutions thanks to crossover and mutation operators adapted heuristically to the problem.

A typical line of research in this area consists of designing new genetic operators adapted heuristically to the problem. For example, in [36] an approximation to the resolution of the job shop scheduling problem is provided by means of a genetic algorithm and an inversion operator with the objective of minimizing the makespan. Another study [37] also propose a resolution method for the job shop scheduling problem based on a new type of genetic algorithm called immune genetic algorithm, IGA. IGA is a variant of genetic algorithms that incorporates an operator additional to crossover and mutation, called immune operator. The immune operator is applied just after crossover and mutation and consists of a double process: vaccine and immune selection. The vaccine process consists of modifying the organisms in accordance with a heuristic criterion based on the specific information of the problem. The immune selection consists of determining in every pair father - son of organisms that one that possesses a major number of antibodies after the vaccine (adaptation to the environment).
In the majority of studies, researchers propose resolution algorithms for very simplified models of industrial plants. However, in [38] relatively complex job shop scheduling problems are considered, with release and due dates and several tardiness reduction objectives, where genetic algorithms combined with search space reduction heuristics are applied to solve the problem.

With respect to flexible job shop scheduling problem, besides the work mentioned above [3], [32], other researchers introduce a genetic algorithm applied to this problem [39], which improves some known strategies combining them to find the best criterion at every step of the algorithm. The initial population is generated by means of the "approximation by localization" strategy, improved with a reordering of work orders and machines and an optimal search in a case-table. The reproductive individuals are selected by the methods of binary tournament, n-ary tournament, and linear hierarchy. POX crossover operator (crossover based on order that preserves precedence) and PPS mutation (mutation by shift that preserves precedence) are used in order to generate feasible solutions. In this line, a different study introduces a genetic algorithm to solve the job shop scheduling problem with alternative routings under four optimization criteria: mean flow time, makespan, maximum lateness and JIT manufacturing [40]. First, the mathematical model of the problem is formulated as a problem of mixed integer linear programming. The aim is to have an exact method of resolution that could be applied to small instances of the problem and that allows to evaluate the quality of the genetic algorithm. This genetic algorithm is standard, with a coding and a crossover technique aimed to obtain feasible solutions.

Finally, the problem of dynamic advanced planning and scheduling, DAPS, is handled in [41]. One of the fundamental aspects of the DAPS problem, in which new orders arrive continuously, is to determine how often it is necessary to re-plan. In the paper, the policy of periodic rescheduling by freezing interval is adopted. This way, the schedule is checked periodically from the point of replanning, but not whenever a new order arrives.

To summarize, the review on genetic algorithms and job shop scheduling offers a global perspective where some interesting elements can be considered in this project, such as problem decomposition, stochastic and heuristic genetic operators, or dynamic scheduling by freezing intervals.

3. GENERAL FRAMEWORK OF THE PROJECT

This research project aims at developing suitable solutions to support collaborative planning and scheduling in complex and dynamic supply chains by providing a software architecture and a methodology to define cooperation in a distributed environment. This research is being undertaken under the frame of the PRORRECO project (Grant PI2008-08, funded by the Bask Government in Spain).

The project contributes to gain the competitive advantage at each single company as well as at the extended enterprise, since it carries out an in-depth analysis of the implications of changes occurred at a certain point of the supply chain for other nodes. This way, the different production schedules of the supply chain are connected in order to optimize the global planning of the supply chain.

The context of the project considers a supply chain composed of different plants having initial production schedules that are dispatched to the shop floors. At this point, the system starts monitoring the execution of the schedules and gathering information about new demand requirements, due date changes, machine failures and other events. Each factory in the supply chain will make use of a capacity model in order to check the feasibility of new production orders or added capacity demand coming from internal exceptions in the factory. Besides, it will communicate with suppliers so as to ensure the availability of raw materials before new demand requirements.

The general framework of the project using a decentralized approach can be defined by means of three subsystems of communication based in agents (see figure 1):

- A communication subsystem inside the plants, which will manage the unforeseen events that may lead to a rescheduling of part or the entire production plan (internal exceptions).
- An inter-plants communication subsystem, which will manage the events produced in a plant that may affect other plants.
- A supply chain communication subsystem, which will manage events, occurred in a plant that can affect suppliers and/or customers.
It is, therefore, essential to establish a very close communication among all the links in the chain. This communication is a two-way flow of information that intends to synchronize the entire supply chain. Thus, every exception or emergency in any node is communicated immediately to the rest of the supply chain so all the members can act accordingly. Each stage in the chain must "march to the same drumbeat" A collaborative work of the chain gets several benefits for the whole chain. There is a growing recognition that effective supply chain management will achieve the twin goals of cost reduction and service enhancement.

As part of this project, a prototype based on these objectives will be developed using an advanced multi-agent architecture. The system architecture is a decentralized architecture where a software agent represents each node of the supply chain. A web-based agent system seems to be a suitable design paradigm in order to enable a dynamic collaboration between different nodes of the supply chain [2]. Each factory will optimize its own jobs locally. But interactions between nodes will be solved through a common supply chain communication subsystem, where events affecting more than one plant will be handled. The selected approach will make use of distributed control, where decisions are made at each node taking into account the information exchanged with other nodes of the supply chain. This means that each plant will be autonomous as regards production planning and scheduling but, at the same time, will take advantage of the common information of the supply chain, so as to allow a dynamic collaboration between the agents.

Regarding internal exceptions, once a production schedule is launched, manufacturing begins. If an exception occurs at a shop floor, the affected operations at the current production schedule will be identified and the feasibility of the solution will be verified. In order to do this, it is absolutely necessary to identify the possible exceptions that the application is going to support and to implement the most suitable algorithms to achieve the best global solution. The aim is to reach a solution that is optimal for the entire supply chain, not for each company individually.

Once an exception occurs, the affected operations at this node will be rescheduled, taking into account the available capacity at the active production schedule. Furthermore, the solution feasibility will be checked externally. In case unplanned demand is introduced in the system and there are not enough raw materials at the warehouse, the affected node will send a material request to one or several suppliers, who
might need to send requests to their own suppliers. Once answers are received, the most optimal one will be selected considering on-time delivery and costs. The solution will be communicated to the customer at the point where the unplanned demand was created.

It is also important to define when a rescheduling process should be accomplished in response to disruptions or changes at the shop floor. At this stage, small deviations between planned start and end times for each job often happen but these are usually ignored, since they will increase the nervousness of the system but will not significantly improve results. This will mean instability and separation from the original schedules, thus leading to increased costs. One important performance measure of the schedule quality must be the distance between schedules, which evaluates the differences between schedules. But large deviations or events affecting the resources capacity must be handled in order to reduce the impact of these disruptions on the schedule. There are two common methods to repair a schedule, namely total rescheduling or partial rescheduling. The second one is considered as more interesting since it brings a solution in a short time and preserves the original schedule as much as possible.

Furthermore, in the case of unplanned demand, the system will employ Available-to-Promise (ATP) and Capable-to-Promise information, so that the marketing department will be able to negotiate with the customer order due-dates, thus providing a basis for real-time order promising [8]. As soon as a new urgent order arrives, ATP will be checked. In case we obtain a positive result, confirmation will be sent to the client. Otherwise, CTP will be checked. If CTP is positive, the first step is to check materials availability. If there is enough material available, the supplier will be notified so that it will be recorded in the historical demand of the company. Otherwise, the next step is to decide whether to order the required material. If the order is finally made but the supplier cannot fulfil it or if the order is not made, the client’s order will not be accomplished and a notification will be sent to the customer, else a rescheduling process will be started taking into account the new order.

5. CONCLUSIONS

This research aims at building a proactive exceptions management solution that operates in real time by monitoring possible disturbances and processing them on a wide basis, thus considering their impact on the overall supply chain. In this way, a suitable synchronization of the different production schedules will be achieved, thus leading to a global competitive advantage over competitors.

In order to do so, most common exceptions must be previously typified and suitable algorithms for optimizing global results in the supply chain must be selected. The target solution must optimize the supply chain and not each node independently.

Literature review showed that genetic algorithms can offer an interesting solution for problem decomposition, stochastic and heuristic genetic operators, or dynamic scheduling by freezing intervals. Furthermore, although SCOR and CPFR are successful models to reduce the bullwhip effect and best meet customer demand, it is not enough to share information but to synchronize the processes in the supply chain.

One basic requirement of the project is to establish a close communication flow inside the supply chain. A two-way information exchange between different nodes of the supply chain is necessary, where exceptions occurred at a certain node will be communicated to other nodes, so as to coordinate work and get feasible and good global solutions. Cooperative scheduling will then lead to synchronized production schedules, increase the supply chain efficiency and better comply with on-time delivery orders.

Some expected benefits of the system are the following: a better information visibility, a fast notification of disturbances throughout the supply chain, a synchronized production and a collaborative production scheduling. These benefits will involve important cost savings and a competitiveness increase for the supply chain as a whole, which is very important in order to survive at the new global business market.

Next stages of the project will include a prototype development and systematic tests that will shed light on the suitability of the architecture and algorithms selected.

REFERENCES


Comprehensive performance analysis of cross-docking system considering transportation consolidation

Yifan Wu¹, Ming Dong¹* and F. Frank Chen²

¹Department of Industrial Engineering and Management
Shanghai Jiao Tong University
Shanghai 200240, P.R. China

²Center for Advanced manufacturing and Lean Systems
The University of Texas at San Antonio
San Antonio, Texas 78249-0670, U.S.A

ABSTRACT

As a key lean logistic strategy leading to dramatic inventory investment and transportation cost reduction, cross-docking is first introduced and applied by Wal-Mart, the retail giant. The successful practical implementation triggers various academic study on this subject, but comprehensive performance analysis of cross-docking system considering inventory holding cost, transportation cost and backorder cost, is hardly existent in the literature to the best of our knowledge.

This paper aims at providing a rapid approximation solution other than simulation to evaluate the performance of cross-docking centers (CDC). Multi product and customer types are included. Each type of product is offered by a single supplier and each type of customer may require different set of items from the CDC, while orders from different types of customers follow Poisson processes with non-identical arriving rates. The replenishment order may not be placed to suppliers until certain amount (i.e. order point) of customer orders are accumulated for the sake of transportation consolidation. When all types of items a specific customer order requires arrive at the CDC they are delivered to the customer immediately. To keep the analysis tractable, a synchronization assumption that no mixing occurs between sets of customer orders, is imposed. Since the items filling on one customer order may not arrive simultaneously, the items arriving earlier are stored in the CDC until the last item arrives, and the stored items make up the inventory. M/G/∞ queueing system is employed to model the replenishment order process, with x denoting the order point for specific supplier. A constant cost of each delivery is considered. Unit inventory and backorder cost is referred as h and b, respectively. The total average cost is then computed.

1. INTRODUCTION

As a key lean logistic strategy leading to dramatic inventory investment and transportation cost reduction, cross-docking is first introduced and applied by Wal-Mart, the retail giant. The successful practical implementation triggers various academic study on this subject, an open fundamental problem is the performance evaluation of cross-docking centers (CDC). However, comprehensive performance analysis of cross-docking system considering inventory holding cost, transportation cost and backorder cost is hardly existent in the literature to the best of our knowledge.

Most research focuses on only one of the factors listed above. The system-wide inventory cost is considered in Waller et al [1] to study the impact of cross-docking on a two-stage system, which consists of a distribution center and several retail stores. The freight moving distance in the CDC is widely used as a surrogate of inventory level for its modeling simplicity. Two solution methods, three-phase heuristic procedure and genetic algorithm, are developed to solve the non-linear mathematical model with the objective of minimizing the travel distance of the pallets in a mail distribution center in Oh et al [2]. A layout problem aimed at reducing travel distance of workers is addressed in Bartholdi and Gue[3]. The best shape of CDC is discussed in Bartholdi and Gue [4] considering just freight travel distance. Vis and Roodbergen [5] models the temporary storage locations problem as a novel application of the minimum cost flow problem, where cost represents the travel distances of the forklift trucks.

Transportation cost is mainly included in a body of literature on shipment consolidation centers. A shipment consolidation problem is modeled using a simulation model and a probabilistic analytical model, respectively in Higginson and Bookbinder [6] and Bookbinder and Higginson [7]. Integer programming formulations are

* Corresponding author: Tel: +86 21 34206101; E-mail: mdong@sjtu.edu.cn
developed in Croxton et al [8] to study the merge-in-transit centers. Furthermore, solution methods for addressing operational issues in these systems are provided.

In this paper, a CDC operating in “pull” strategy is modeled with queueing theory. Multi product and customer types are included. Each type of product is offered by a single supplier and each type of customer may require different set of items from the CDC, while orders from different types of customers follow Poisson processes with non-identical arriving rates. The replenishment order may not be placed to suppliers until certain amount (i.e. order point) of customer orders are accumulated for the sake of transportation consolidation. Backorder cost accumulates until order is filled. When all types of items a specific customer order requires arrive at the CDC, they are delivered to the customer immediately. To keep the analysis tractable, a synchronization assumption that no mixing occurs between sets of customer orders, is imposed. Since the items filling one customer order may arrive simultaneously, the items arriving earlier are stored in the CDC until the last required item arrives, and the stored items make up the inventory.

The reminder of the paper is organized as follows: Section 2 provides a brief description of the analytical model. The product replenishment process is modeled using \( M/G^\infty \) queueing system in Section 3. Section 4 derives expressions of three main types of cost considered in this paper. Conclusions are drawn in Section 5.

2. MODEL DESCRIPTION

Let \( P = \{1, 2, \ldots, m\} \) denote the set of all product indices. For any subset of product \( K \subseteq P \), we say an order is of type \( K \) if it consists of one unit of each product in \( K \) and zero units in \( P \setminus K \). Customer orders of type \( K \) arrive at the system following a stationary Poisson process, denoted \( \{A_K(t), t \geq 0\} \), with rate \( \lambda_K \), see Figure 1. It is assumed that each order’s type is independent of the other orders’ types and of all other events. \( D \) is used to denote the set of all order types.

![Figure 1: Cross-docking center](image)

For each product \( i \), let \( D_i \) denote the family of subsets of \( D \) that contains \( i \). Then it is clear that the demand process for product \( i \) is also a Poisson process with rate:

\[
\lambda_i = \sum_{K \in D_i} \lambda_K
\]  

Orders are filled on a FCFS basis. Upon the arrival of any customer order, the customer order is transformed into several product orders according to the subset of products this customer order requires. Replenishment order for a specific type of product will not be placed until a certain amount (order point) of product orders for this type of product is accumulated. Unfilled order is backlogged and is charged a cost rate of \( b \). Once the suppliers receive the replenishment order, the amount of products required will be delivered to the CDC in a truck. The transportation cost consists of constant cost, \( C \) and variable cost related to the product amount. The variable cost does not vary when different order points are chosen, thus only constant transportation cost is included in total cost.

Since the product demand rates, \( \lambda_i \), and lead-times, \( l_o \), may differ in product types, items filling one customer order may not arrive simultaneously. The items arriving earlier are stored in the CDC until all the items required arrive, and the stored items make up the inventory. Inventory holding cost rate is \( h \) per item. Moreover, the product unloading and loading time is assumed to be negligible. Thus, the order is immediately satisfied once the trucks arrive at the CDC.
3. **PRODUCT ORDER QUEUE**

It is obvious that each product order queue can be modeled by an $M/G^X/\infty$ queueing system. For any product type $i$, let the order point be $x_i$. The waiting product order queue is markovian and the state space is $\{0, 1, 2, \ldots, x_i - 2, x_i - 1\}$, then the transition rate matrix $Q$ is given by

$$Q = \begin{pmatrix}
-\lambda & \lambda & 0 & 0 & \cdots & 0 \\
0 & -\lambda & \lambda & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & -\lambda & \lambda \\
\lambda & 0 & 0 & 0 & 0 & -\lambda
\end{pmatrix}. \quad (2)$$

The balance equations are:

$$\begin{align*}
    p_0(-\lambda) + p_{x_i-2}\lambda &= 0 \\
p_i\lambda + p_0(-\lambda) &= 0 \\
p_i\lambda + p_2(-\lambda) &= 0 \quad (3) \\
\vdots \\
p_{x_i-2}\lambda + p_{x_i-1}(-\lambda) &= 0
\end{align*}$$

from which, together with $\sum_{j=0}^{x_i-1} p_j = 1$, follows that

$$p_j = \frac{1}{x_i}, \quad (4)$$

where $p_j$ is the limiting distribution probability, $j=0,1,\ldots,x_i-2, x_i-1$.

4. **COST ANALYSIS**

4.1. **CUSTOMER ORDER WAITING TIME AND RELATED COST**

For a customer order of type $K$, when it arrives at the system, it sees the state of the product order queues with probability $p_j = 1/x_i$, where $i \in K$, $j=0,1,\ldots,x_i-2, x_i-1$. The waiting time of the customer order $K$ for product $i$ is:

$$W^K_i = n e(\lambda_i) + l_i, \text{ with probability of } \frac{1}{x_i} \quad (5)$$

where $e(\lambda_i)$ is an exponential random variable with mean equals to $\lambda_i$, $n=0,1,\ldots,x_i-1$.

Clearly, different sets of decision variables $(x_1,x_2,\ldots,x_M)$ result in different customer order waiting time, and the waiting time of customer order $K$ can be expressed as:

$$W^K = \max_{i \in K} (W^K_i) \quad (6)$$
In order to calculate customer order waiting time, customer order waiting time for each type of product $i$ is characterized using CMT families of distributions with two parameters, i.e.

$$W^K_i \sim \text{CMT}^{2^{\nu}}(\alpha_i, \beta_i) \tag{7}$$

Then the customer order waiting time takes the form

$$W^K = \max_{i \in K} (W^K_i) \sim \text{CMT}^{2^{\nu}}(\sum_{i \in K} \alpha_i, \sum_{i \in K} \beta_i) \tag{8}$$

The expected backorder cost for customer order $K$ is

$$E(BC_K) = \lambda_K b E(W^K) \tag{9}$$

The following approximations are used to derive the parameters $\alpha_i$ and $\beta_i$:

$$E(W) \approx \sqrt{\frac{u^2 + v^2}{2uv}} \ln[\exp(\sqrt{\frac{2u(\gamma + \ln \alpha)}{u^2 + v^2}}) + \exp(\sqrt{\frac{2u(\gamma + \ln \beta)}{u^2 + v^2}})]$$

$$E(W^2) \approx \frac{\pi^2 + 6(\gamma + \ln \alpha)^2}{6u^2} \Phi\left(\frac{\sqrt{6} \ln(\alpha^{\nu} / \beta^{\nu})}{\pi \sqrt{u^2 + v^2}}\right) + \frac{\pi^2 + 6(\gamma + \ln \beta)^2}{6v^2} \Phi\left(\frac{\sqrt{6} \ln(\beta^{\nu} / \alpha^{\nu})}{\pi \sqrt{u^2 + v^2}}\right)$$

$$+ \frac{\pi \ln(\alpha^{\nu} \beta^{\nu}) \sqrt{u^2 + v^2}}{6u^2 v^2} \phi\left(\frac{\sqrt{6} \ln(\alpha^{\nu} / \beta^{\nu})}{\pi \sqrt{u^2 + v^2}}\right)$$

where $\nu \approx 0.5772$ is Euler’s constant, $W \sim \text{CMT}^{2^{\nu}}(\alpha, \beta), \Phi$ and $\phi$ denote the cdf and pdf of a standard normal random variable, respectively.

Combine equations (8) and (10), the expected backorder cost caused by type $K$ customer order can be obtained. Thus, the total expected backorder cost is

$$E(TBC) = \sum_{K \in D} E(BC_K) \tag{1}$$

### 4.2. Inventory Holding Cost and Transportation Cost

Since part of the items that an order requires may arrive earlier than the others, these items construct majority of the inventory in a CDC. And the inventory cost referred to type $K$ customer order can be calculated as follow.

$$E(IC^K) = E\left(\sum_{i \in K} h(W^K_i - W^K_i)\right) = h \sum_{i \in K} \left( E(W^K_i) - E(W^K_i) \right) \tag{2}$$

As $E(W^K)$ and $E(W^K_i)$ are available from section 4.1, equation (12) can be applied directly. And the expected total inventory cost is

$$E(TIC) = E\left(\sum_{K \in D} IC^K\right) = h \sum_{K \in D} \sum_{i \in K} \left( E(W^K_i) - E(W^K_i) \right) \tag{3}$$

The expected total transportation cost can be expressed as

$$E(TTC) = C \sum_{i \in \mathbb{P}} \lambda_i / x_i \tag{4}$$
The expected total cost is the sum of the all three types of cost:

\[
E(TC) = E(TBC) + E(TIC) + E(TTC)
\]

\[
= \sum_{K \in D} \lambda_K bE(W^K) + h \sum_{K \in D \cap i \in K} (E(W^K) - E(W^K_i)) + C \sum_{i \in P} \lambda_i / x_i
\]  

5. CONCLUSIONS

This paper provides a comprehensive performance analysis of cross-docking center operating under a “pull” strategy. As cross-docking technology is regarded as a lean logistics strategy which maintains high customer service levels while reducing inventory level at the CDC to the minimum. Transportation cost saving is another main advantage achieved by freight consolidation at the CDC. All things considered, inventory level, customer response time and transportation cost are selected as comprehensive performance measures and are transformed to total cost. A fast approximation method other than simulation is developed to evaluate the system performance. This method can be used to analyze the impact of the operational issues, mainly product order points, on the system performance. It also provides a solid foundation for the system optimization, which is a main future research direction.

ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China (70871077) and The National High Technology Research and Development Program of China (2008AA04Z104).

REFERENCES


A System Dynamics Based Evaluation of Collaboration Strategies to Manage Demand Gap in New High-tech Product Diffusion

Xumei Yuan\textsuperscript{a,}*, Ling Shen\textsuperscript{a†}, Jalal Ashayeri\textsuperscript{b}

\begin{itemize}
  \item \textsuperscript{a}Department of Industrial Engineering
  \textsuperscript{b}Dept. of Econometrics & Operations Research
  \textsuperscript{a†}Teacher at E&A College, Hebei Normal University of Science & Technology, Qinhuangdao, China
  \textsuperscript{b}Tilburg University
  \textsuperscript{a}Yanshan University
  \textsuperscript{b}Tilburg, The Netherlands
\end{itemize}

\textbf{ABSTRACT}

As companies decide where to allocate their new product and capacity development budgets in a tough economy, they need to focus on investments that will bring the most value—during a downturn and beyond. The level of engagement of customers downstream is often cited as the number-one factor in its ability to survive a downturn. Given this, companies require to evaluate collaboration strategies to properly manage the process of new product and capacity developments. There's no question that many high-tech manufacturing companies operate in a context of high process and market uncertainties due to shorter product life cycles. When introducing a new product, these companies must manage the cost of supply, including the cost of capacity and inventories, with revenues from the product’s demand over its life cycle. However, in early phase of introduction after earlier buyers purchase, there might be a demand gap for a period followed by a sudden surge. An important decision when launching a new product is the sizing of the capacity but to stay responsive and serve the market downstream after such gaps, collaboration is also a necessity. It is the intention of this paper to show that the chosen level of capacity does not effect significantly on the dynamic of gap in the demand trajectory, while collaboration does help managing an extensive gap in a new high-tech product diffusion dynamics. We study the impact of different collaboration strategies like Vendor Managed Inventory (VMI), Jointly Managed Inventory (JMI), and a Collaborative Planning, Forecasting & Replenishment (CPFR) model using system dynamics based simulation and compare the results with a non-collaborative chain. Our results yield insights into effectiveness of collaboration in managing the dynamics of demand gap.

1. INTRODUCTION

With business conditions steadily getting tougher, companies have to be more cautious with the introduction of new products. While managing the limited budget, they must stay tuned and responsive to the customers’ needs. All high-tech products have certain short life cycles. During this period significant changes are made in the product demand behavior in the market. Since an increase in profits is the major goal of any company that introduces a product into a market, understanding the demand gaps in product’s life cycle is very important. Some companies use strategic capacity planning and others follow the basic rules of the different life cycle phases and decide about capacity, and adjust it later.

The shape of traditional product life cycle usually has rather short introduction time, quickly growing speed in the products’ growth stage, long duration of the products’ maturing stage and declining stage. And the declining speed is quite slow.

* Corresponding author: Tel.: 0086-335-8387519; Fax: 0086-335-8057035; E-mail: yxm@ysu.edu.cn
† Teacher at E&A College, Hebei Normal University of Science & Technology, Qinhuangdao, China
As early as 1950’s, the basic concepts of Product Lifecycle Management (PLM) are proposed by Dean [1] and Levir [2] in the field of economy management, with the purpose of studying the products’ market strategy. According to the evolutionary process in the market, they divided the product life cycle into four stages i.e. the introduction stage, the growth stage, the maturing stage and the declining stage (see Figure 1, left). With the technological development over 50 years, the concept and the connotation of product life cycle are unceasingly changing. Now that the classical product life cycle has been widely applied, the new technological development has changed the traditional shape of common product life cycle as shown in Figure 1, right.

![Figure 1 – Classical (left) versus High-Tech (right) Product life cycle curves.](image)

Traditionally, the sales volume slowly increases in introduction stage while the growth of sales is rapidly rising in the growth stage. Once the growth of products’ sales becomes slower, it indicates that the products are entering into maturing stage, in which stage the sales will reach the peak. Afterwards, their sales tends to descend and decreases more and more rapidly after entering the declining stage until the products completely withdraw from the market, completing its entire process of life cycle [3]. The supply chain point of view in the first one should carefully observe the demand and react to it as fast as possible. This requires supply chain responsiveness. In the third stage the objective is to replenish the stock most economically, requiring leanness from the supply chain. The last stage is making decisions about the phase out quantities to be produced and the price corrections.

As is shown in Figure 1, compared with the traditional products, the shape of product life cycle of high-tech enterprises is extremely different with the following main characteristics:

1. Products’ research and development cycle is comparatively long, mainly because of the products’ complexity and the high technical content. The R&D cycle of the fundamental innovation can be especially longer.

2. The introduction stage and the growth stage would witness a sharp increase in demand if the product gains acceptance. The body of potential customer needs to go through a “silence time” to get know, accept the use, and the benefit of the new product.

3. As a result of the rapid advancements in technology, cut-throat competition, the wide range proliferation of high-tech products, and the technical overflow effect, products’ maturing and declining stages are shortened and merged into one stage, which put immense pressure on the high-tech companies to constantly weed out the old product and bring in a new product in the market. New product introduction has a huge impact on the length of stage three (decline stage). Usually certain high-tech enterprises which have formed the monopoly status, for the purpose of tapping the maximum market potential, create the demand diligently and eliminate the old products only and when necessary, extending decline stage. However, generally compared with the traditional products, products’ declining speed is faster.

There are still two kinds of related high-tech product life cycle theory besides product life cycle characteristics proposed above in high-tech industry. First is “the demand gap” during the silence time [4]. Usually after initial promotion of a high-tech product, demand comes from some early buyers, mainly technically literate customer base or very curious consumers. This introduces a surge but after sometimes then sales hit a lull as they try to move on to mainstream buyers, which is totally beyond expectation. This creates a sales “gap”, see Figure 2. The gap exists because mainstream buyers are not decided yet. These customers have different buying criteria in the next group up in the demand chain. The mainstream consumers are willing to wait and study further the benefits of the new technology. They purchase only after gaining enough confidence in the products. These new buying requirements are
fundamentally different from those of the early buyers and only by careful reading/reacting one can improve demand position and move forward without going out of business in the process. The population size of this kind of customers is large; therefore if they decide to buy there will be a huge “surge” in the product demand after the gap. This second effect is due to sudden reappearance of demand at large quantity. An unfortunate reality is that this mainstream group may opt not to buy the product and such the demand would never raise again.

Many high-tech products do not meet the basic conditions of the traditional products in market; moreover many companies producing products of this nature are not vertically integrated and cannot offer alone a solution when they promote the new products. This complicates the problem further.

The classical life cycle generally assumes normal distribution (bell-shaped) curve, while the life cycle curve of high-tech product assumes the S shape curve distribution. After the high-tech products enter the market, following a short introduction time, they enter the growth period quickly, and the growth of the demand is rapid in the growth period and then the high-tech products enters the declining period. Because of the impacting of new products' emerging one after another incessantly, the customer demand shifts and reduces constantly.

With the above discussion we believe that high-tech products with the fundamental innovation require an end-to-end (E2E) process analysis from products’ primary development phase to the products’ launching up to customers' approval. However, an E2E analysis to study unexpected gaps, manage the demand behavior, minimize stockouts, increase sales and profit within a disintegrated chain is a rather difficult task and requires collaboration among different entities. Given this fact in most of the high-tech supply chains, each entity focuses on capacity management and agility within own entity in order to address demand surge after the gap. Therefore, this study aims at comparing a non-collaborative chain with a collaborative one using system dynamics based simulation. We show that collaboration can assist managing the dynamics of demand gap in the introduction of a new high tech product.

The next section reviews the relevant collaboration strategies to be simulated. The third section presents the system dynamic causal models that are simulated. The fourth section introduces the simulation results, analyzes them, and derives some answers for demand gap management. We conclude with discussion of our main results, managerial and theoretical implications, and directions for future research in the fifth section.

2. COLLABORATION STRATEGIES

2.1 BACKGROUND ASSUMPTIONS

The theoretical high-tech company's supply chain is given in Figure 3. In order to have realistic data, we simulate the life cycle model of an innovative high technology product, considering products like computer or mobile phones. The cycle is assumed to be short, less than two years (100 weeks) and we assume that demand is a low constant during the “gap effect” period. At the end of the “gap effect” time, the product demand steps up, grows with a given rate achieving its peak. Then the demand slowly comes down entering the declining stage. Hereafter, the decline rate increases gradually until the product is withdrawn from the market.

Planning of short life cycle products is so difficult because the rate of demand growth and decline is rapid. Each stage, from ‘Merchandise Test’ to ‘Early Read’, then to ‘Replenishment’ and finally to ‘End of Life’ has differing supply-chain parameters. Quite often, a model developed for replenishment does not capture the chain realities when a product is nearing end of life. Also it is difficult to create a forecast from a merchandise test and know how to scale that
to a chain and then use it to plan the early demand. Clearly we deal with multiple problems and as such an E2E study is necessary.

Here below we briefly discuss the four conceptual models to be simulated.

2.2 Traditional Supply Chain Stock Management Model

In the traditional supply chain, the company at each level is an independent entity, as shown in Figure 3.

According to the Figure 3, the products, which exist as raw material, are firstly shipped from the suppliers’ to the manufacturer site, and the manufacturer processes or produces products and then ships them to the distributor in large volume. And the distributor further reduces the product batch according to next entity (retailer) order and ships them to each retailer. Retailers face the end user and obtain the most accurate information of sales. According to the sales data, the retailers determine their own safety stock, order size, and place their order to the distributors. Similarly the distributor forecast the demand, determine safety stock adjustment needed, calculate the order quantity and place it to manufacturer. And manufacturer does the same. The classical bull’s effect shown in “beergame” can be observed here even when there is no demand gap. However, when there is such a gap due to introduction of a new high-tech product massive product stock is built up the stream or extra capacity reserve must be provided in order to react to the demand surge after the gap. As a result, vast amount of funds in the chain is wasted. Therefore, the policy-makers of the supply chain are seeking new collaborative management models to reduce the costs and maintain the supply chain agility.

2.3 Vendor Managed Inventory Model

VMI (Vendor Managed Inventory), is the simplest cooperative strategy between the customers and the suppliers. With the thought of integrated supply chain management, it replaces the traditional inventory management model and enables the entire supply chain minimizing the stocks. The main goal is to provide a mutually beneficial relationship where both sides will be able to more smoothly and accurately control the availability and flow of goods.

Figure 4 shows that, compared with the traditional supply chain, the final customers’ demand information is not only available to the retailers but also to the distributor. Under a frame of bilateral agreement, the distributor manages the inventory of the retailer. The distributor has access to the retailer’s inventory data and is responsible for generating purchase orders. This proactive replenishment realizes a rapid reaction to the changes in the customer demand. The VMI can be also that , the supplier manages the downstream manufacturer’s inventory , or the manufacturer manages the distributor’s inventory. But in this paper, we only consider the situation in which the distributor manages the retailer’s inventory.

VMI implementation improves service level which in turn attracts more customers to the retailers but also reduces several costs like inventory carried over costs, obsolescence cost which is so important in high-tech supply chains due to short life cycles of products, transportation and warehousing costs, and transaction costs. The implementation would establish the foundation of full trust, and ultimately in long run enhance the strategic partnership [5].
2.4 **Joint Managed Inventory Model**

Joint Managed Inventory (JMI) enhances the synchronizing bound of the supply chain through a coordinated and a harmonious mechanism between the supply chain members in order to solve the problem caused by the demand fluctuations. It’s a kind of inventory management based on risk sharing.

In the general supply chain inventory management, each member in the chain keeps safety stocks for both independent demand and dependent demand. The general format of JMI is shown in Figure 5. The basic thought behind it is to reduce or eliminate distorted information on demand. With the objective of reducing the influence of the bullwhip effect as much as possible, it transforms some parts of independent demand inventory into a joint one, and as a result to reduce the level of the safety stock and the average stock in the supply chain.

JMI is a kind of consignment stock management method with a coordination center. This method centrally embodies the following: locate a number of stock control positions in the supply chain which are controlled centrally by the involved parties, for example raw materials, partly completed, manufactured stock (see Figure 5). This enhances agility and responsiveness while solves “the phenomenon of demand oscillation” in the supply chain system. Joint Managed Inventory requires some level of cooperative relationship, and it is a kind of risk sharing stock management model.

![Figure 5 – JMI supply chain model](image)

In Figure 5, there are joint managed inventory between the supplier and manufacturer, and also between the manufacturer and distributor. Thus, the manufacturer can face directly the customer demand of the retailer, and as the inventory is managed together, the risk is shared.

2.5 **Coordination Supply Chain Management Model**

Coordinated supply chain management follows the idea of Collaborative Planning, Forecasting and Replenishment (CPFR) through the development and extensive application of the information technology in the supply chain, in order to reduce the cost of the supply chain and enhance its operating efficiency.

CPFR is not only one idea, but also a series of activities and the process, making it possible to expand sales and reduce stock and achieve high service levels [6]. Figure 6 shows that, compared with the traditional supply chain, the customer’s final demand information is directly communicated to all entities, and therefore each entity is able to reduce unnecessary stocks, maintaining the stock at a low level. Thus the entire supply chain stock reduces. As for planning, sharing information would allow each entity to accurately grasp the capacity limits of the next entity and the final customer demand fulfillment level, thus correctly set the stock levels down the stream.

![Figure 6 – The CPFR supply chain model](image)
3. System Dynamics Models

Here we assume that raw material supply capacity is sufficient and as such considers a three-level chain containing the retailers, the distributor and the manufacturer.

3.1 Traditional Model Casual Loop

In the traditional model, each entity receives orders from downstream, determines its own stock adjustment rate by forecasting, and then places order to the upstream entity. In this case, only retailers are facing the real customer demand. The chain performance is usually dependent on the accurate sizing of orders transmitted upstream and the upstream capacity to fulfill the orders on time. Figure 7 shows the causal loop of the traditional high-tech supply chain.

![Figure 7 – Causal loop diagram of traditional model](image)

3.2 VMI Casual Loop

VMI is a kind of cooperative strategy enabling each link of the enterprise to obtain the least inventory cost. VMI not only reduces the enterprise cost of each link in the supply chain, but also overcomes the information distortion problem. In the case under study, the retailer stock quantity is controlled by the distributor. In this model, in order to monitor the inventory position and to compare the results with the traditional chain, we still call it “retailer stock”, but it is managed by the distributor. When the retailers face the final customer demand, they transmit the information from point of sale (POS) via right technology like EDI to the distributor (see Figure 8). Then distributor delivers goods to adjust the retailers’ inventory, and orders goods from manufacturer based on a new forecast. When the retailers receive the goods, the order quantity is decided by the distributor in order to satisfy retailers’ stock level target.

3.3 JMI Casual Loop

We consider the following JMI situation. The retailer faces the final customer demand directly, adjusts its inventory by forecasting. The distributor and the manufacturer manage inventory together, sharing the risk. Figure 9 shows the casual loop in which there is a joint inventory instead of manufacturer inventory and distributor inventory separately.

As the stock lies with both manufacturer and distributor, the transportation cost/distance between the retailers and distributor may increase. The hypothetical entity manages all the resources involved in this joint stock through the highly effective information transmission, and enables the distribution of the good to avoid any delay through close cooperation. Ultimately, this approach may create the condition for the realization of a lean supply chain, with minimal stock and on time order delivery.
3.4 CPFR CASUAL LOOP

The purpose of the CPFR is to remove “the bullwhip effect”, and to realize a truly collaborative supply chain. By implementing CPFR, the retailer, the transportation business, the third party physical distribution agent, the distributor and the manufacturer coordinate activities from the production operation plans to the execution. CPFR pays great attention to the sharing of customer demand information among partners to enhance the efficiency of the supply chain.

Under CPFR management model, various entities share the information on demand, and adjust their own target inventory. The causal loop reflects this major change compared with the traditional chain (see Figure 10).
4. RESULTS AND DISCUSSION

The four supply chain situations discussed above are implemented in Vensim, a system dynamics based simulation software. Several simulation scenarios are considered here focusing on the stock levels in the supply chain given the high-tech product demand. There are many others runs that cannot be covered in this paper due to the limited space. In total, 10 scenarios are simulated with different parameter settings namely, one traditional supply chain as base reference, two VMI scenarios, four JMI scenarios, and three CPFR scenarios. We skip detailing these information and pay attention to the highlights of all simulation runs. Note all simulation runs were executed for 100 weeks, with an initialization period.

In general, the results uncover a less readily apparent relationship: the manufacturer's share of the resulting profit decreases as the distributor becomes increasingly conservative towards risk, i.e., with smaller risk sharing parameter value. In summary, the results indicate that the manufacturer of a high-tech product would be much better-off increasing the share of resulting profit – to coordinate with the distributor when demand is not very price-sensitive and/or the distributor is increasingly risk-averse. Concerning the stock levels in the chain, each case is producing an outcome that is briefly discussed below.

Traditional supply chain management model operates through a step-by-step forecast, ordering from downstream to upstream and delivering goods from upstream to downstream. In this step-by-step information transmission, the information distortion creates the “bullwhip effect” and that can be easily seen in Figure 11. The “gap effect” in the demand only delays the oscillations. Other tests also show that capacity increase only dampens the bullwhip effect depending on its magnitude but never removes it. Now we can verify whether the other collaborative approaches would improve this result.

Compared with the traditional high-tech supply chain model, VMI shares the customer demand information obtained by the retailer with the distributor, at the same time, the retailers’ stock is controlled by the distributor through the agreement, and the stocks at distributor and retailers are simultaneously adjusted by the distributor, allowing to increase the shipment when necessary. Given more accurate information sharing, we can see the stock levels of distributor and retailers reduced dramatically (see Figure 12), while the manufacture suffers still from stock fluctuation similar to the traditional chain model.
A System Dynamics Based Evaluation of Collaboration Strategies to Manage Demand Gap in New High-tech Product Diffusion

Figure 11-a – Demand curves in the traditional model
Figure 11-b – Change stock levels in the traditional model

Figure 12-a – Demand curves in the VMI model
Figure 12-b – Change stock levels in the VMI model

Figure 13-a – Demand curves in the JMI model
Figure 13-b – Change stock levels in the JMI model

JMI stock management strategy performs significantly better than the previous two situations because it enables the manufacturer and the distributor to control their stocks together, in this way, when facing the retailer's orders changes, they can adjust quickly the stock and that in turn reduces the adjustment cycle for the jointly managed stocks. Therefore, the stock adjustment cycle of the retailers may lag a bit, but the fluctuation margin of the entire supply chain stock reduces. Figure 13 shows that the over performance of JMI is better than VMI.

The CPFR collaboration strategy is expected to produce the best and most reasonable results compared with other approaches (see Figure 14). However, it is close to JMI. As can be noted from the figure despite minor stock variability, the difference the high-tech supply chain is capable of matching the demand after the “gap”.

While JMI performs better than VMI, the choice between JMI and CPFR depends on several other factors. These can range from information technology available to the level of trust existing among the partner entities in the chain, to the format and type of financing system, to the geographical spread of retailers. Therefore, an E2E simulation like this one will not provide the definitive answer but sheds light on the performance of various collaboration strategies in order to synthesize further these strategies and make the final choice.
5. CONCLUSIONS

In this paper the intention was to show that collaborative strategies can reduce the risk of going out of stock in a high-tech supply chain system, when the demand surges after a “gap” time. We compared three strategies with a traditional chain namely: VMI, JMI, CPFR. We showed that the design of an E2E model capturing the main dynamics of demand, can assist selecting the right collaboration strategy. We used system dynamics based simulation modeling based on the cause and effect relational graph, and implemented the models in system dynamics software Vensim.

With E2E modeling like the one presented here it is possible to study and compare the impacts of different collaboration scenarios on the supply chain performance. The paper skipped the description of all scenarios details and directly referred to the main results. The anticipated result was that CPFR performs better than other strategies but results were close to JMI and final choice requires investments analyses over the entire product life cycle. Therefore, it is necessary to perform a detailed financial assessment of the viability of each scenario. The striking and unforeseen results come from profit generated due to level of risk (profit) sharing. It appears that the more trust in taking risk jointly, the more profit is generated for the entire chain. Therefore, collaboration at level of information sharing improves the performance but not necessarily maximizes it, establishing trust and risk/profit sharing is needed.

While some supply chain systems are working towards a demand-driven system by establishing intelligence to quickly “sense and react” to actual demand, some others are trying to “shape” demand based on permanent visibility of market activity. However, in the high-tech environment, shaping the market or adding extra intelligence in early phase of product introduction is too complicated and the supply chain must rely on setting the right collaborative strategy in order to manage the “gap” effect.

The paper did not study hybrid collaborative situations, e.g. part of the chain is operating under VMI and other parts under JMI. Not necessarily the entire chain should follow one unique collaboration strategy. The results obtained thus far suggest the following topics for further study: (a) scenario analysis of the best capacity structure under each collaboration strategy. (b) Extension of the model to include adjusting capacity during the product life-cycle. (c) Extension of the model to include suppliers with limited capacity. (d) Introduction of multiple products.

ACKNOWLEDGEMENTS

The authors would like to thank Scientific Research Fund of Hebei Provincial Education Department and Natural Science Foundation of Hebei Province for supporting this work.

REFERENCES

Joint Replenishment and Supplier Selection using Particle Swarm Optimization

Y.-C. Wang * and W.-J. Li
Department of Industrial Engineering and Systems Management
Feng Chia University
Taichung, Taiwan

ABSTRACT
The classical joint replenishment problem (JRP) deals with the issue of determining a replenishment policy that minimizes the total cost of ordering multiple items from a single supplier. However, in a real supply-demand system, the downstream companies always look for more than one supplier to get a better deal since competition among the suppliers may result in lower price for the items supplied. In this study, we consider the JRP under the condition of sourcing items from multiple suppliers. The major objectives are two folds: selection of proper suppliers for a certain items and determination of the replenishment cycles for them. A particle swarm optimization (PSO) approach integrated with the RAND method, a heuristic that has been proven capable of finding almost as good as optimal solutions for the single-supplier JRP, is proposed in this paper. An illustrative example is then presented to implement the proposed approach to solve the multi-supplier joint replenishment problem.

1 INTRODUCTION
In the past few decades the joint replenishment problems (JRP) have received a lot of attention in that the JRP not only can be viewed a multi-item inventory problem but its concepts can also be applied to lot sizing problems in manufacturing applications. In practice, grouping items into the same order is not uncommon because it may provide some benefits such as savings on ordering costs, on unit transportation costs, and even on purchasing costs, such as when a discount is offered by the vendor if the order amount is greater than a certain quantity. The total cost considered in inventory problems usually consists of three components: the ordering costs, the costs of the ordered items, and the inventory holding costs. Since an order may consists of multiple items in JRP, the ordering costs are composed of a major ordering cost which is independent from the ordered items and the quantity and a minor ordering cost which depends on the ordered items and the quantity in the order. The basic idea of the JRP is to save costs by coordinating items when making replenishment decisions, since the items ordered in the same order share the major ordering cost.

Based on the common assumptions of the traditional JRPs, the items considered in such problems are sourced from a single supplier. However, a downstream company in a supply-demand system always looks for more vendors for a single item since sourcing components from a single supplier is too risky [1]. In addition, competition among the suppliers could offer a lower price and better quality products purchased. In this study, we extend the classical JRP into a multi-item, multi-supplier inventory model, in which the buyer no longer orders the items from a single supplier. Each supplier is capable of offering all the items with different minor ordering cost. Therefore the aim of our research is to develop a particle swarm optimization (PSO) approach for selecting proper supplier for offering a certain items and finding the replenishment policy that incurs a lower total cost.

2 LITERATURE REVIEW

2.1 JOINT REPLENISHMENT PROBLEM

* Corresponding author: Tel.: (886) 4-24517250 ext.3637; E-mail: wangyc@fcu.edu.tw
There have been many heuristic approaches proposed for solving the single-supplier JRPs [2]. These approaches can be classified into two different strategies: direct grouping strategy and indirect grouping strategy. With the direct grouping strategy, items are divided into a predetermined number of groups, and the items within the same group are jointly replenished. An individual cycle time is determined for each group. Each item in the group is replenished every group cycle time. In contrast to the direct grouping strategy, the indirect grouping strategy uses a basic cycle time for each order, although not all products are necessarily ordered every basic cycle time. Some items might have a replenishment quantity to last for exactly one basic cycle time period, but others might have a replenishment quantity to last for several basic cycle time periods. The main difference between the direct grouping strategy and the indirect grouping strategy is that the replenishment cycles for groups in the direct grouping strategy are independent, while the cycles determined by indirect grouping strategy are multiple integers of some basic cycle time. A detailed comparison between the class of indirect grouping and direct grouping strategies was made by Van Eijs et al. [3] as well as Olsen [4]. To solve the classic JRPs, Kaspi and Rosenblatt [5] proposed a heuristic algorithm, called the RAND. In their study of testing over 70,000 randomly generated cases, the RAND method performed better than all the not enumerative algorithms. For those cases where the RAND did not find the optimal solutions, the solutions the RAND obtained were on average within 0.002% from the optimal.

Khouja et al. [6] compared the RAND method with their proposed genetic algorithm. They pointed out that their genetic algorithm has the ability of dealing with constrained problems, but their experimental results revealed that their GA-based method didn’t outperform the RAND in terms of total cost measurement. Olsen [4] proposed an evolutionary algorithm using direct grouping strategy to solve the JRPs. Their proposed algorithm provided lower total cost than the RAND in only 31 of the total 243 different problem settings. Wang and Cheng [7] demonstrated that the RAND is capable of giving a good replenishment plan even under a certain level of uncertainty on the unit holding costs and the customer demands. All the previous research work really indicates that the RAND method is very robust for finding as good as the optimal solution with very little computation effort even with some information uncertain.

2.2 PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) which was initially proposed by Kennedy and Eberhart [8], is one of the latest evolutionary optimization techniques. Its concept was originally from the biological behavior like flocking of birds [9]. The position of each individual particle is represented a solution, the particles fly around the multidimensional space searching for the optimal solution with a moving velocity that is dynamically adjusted based on its own moving experience and its colleagues. In recent years, PSO has been widely applied to various applications. Salman et al. [10] developed a PSO algorithm for task assignment problem. Andrés and Lozano [11] solved the part-machine grouping problem with a certain of restrictions by PSO. Their conclusion revealed that the PSO algorithm could find the optimal solutions on almost all the instances. Mohemmed et al. [12] used PSO to solve the shortest path problem. Their PSO also successfully found near-optimal paths.

Tsai and Yeh [13] presented a PSO algorithm for inventory classification problems with four objectives including minimizing cost, maximizing inventory turnover ratios, maximizing inventory demand correlation, and the combination of the above three. Their algorithm was compared with other classification techniques such as the ABC classification method. They found that their PSO algorithm outperformed the other classification methods commonly used in practice.

3 THE MULTI-SUPPLIER JOINT REPLENISHMENT PROBLEM

From viewpoint of a buyer in a supply-demand system, sourcing components from a single supplier is too risky and gives the supplier too much power in the relationship. Very few JRP studies in literature have tackled the multiple supplier issue. Moon et al. [14] considered a JRP, in which items are procured from multiple suppliers offering quantity discount. They proposed a GA-based approach to solve the problem. In their study, each supplier had their own minor ordering cost and unit cost depending on the order quantity for each item. Their approach determines which supplier the item is sourced from and its replenishment schedule. In their model, however, the ordering cycles for all the items follow a common cycle time no matter if the items are sourced from different suppliers. In practice, the locations of the suppliers may far from each other and even some suppliers are located in different countries, thus a common replenishment cycle for all the items without considering the suppliers does not do anything any good. In this research, we extend the JRP model of Moon et al. [14] by taking account that the common ordering cycles of the items in JRP can be different for various suppliers.

To present the model of this multi-supplier JRP, the following assumptions are made:
1. The demand rate for each item is deterministic and constant.
2. The replenishment lead time is of a known duration.
3. Shortage is not allowed.
4. The unit holding cost for each item type is known and constant.
5. The different suppliers offer different minor ordering costs for each item.
6. Each item can be purchased from only one of a few suppliers.

The following notations are used in this model:

- \( M \): number of suppliers
- \( N \): number of items
- \( j \): index of supplier, \( j = 1, \ldots, M \)
- \( i \): index of item, \( i = 1, \ldots, N \)
- \( s_j \): the \( j \)th supplier, \( j = 1, \ldots, M \)
- \( D_i \): demand rate of item \( i \)
- \( A \): major ordering cost
- \( a_{ij} \): minor ordering cost of item \( i \) that purchased from the \( j \)th supplier
- \( h_i \): the holding cost for item \( i \)
- \( T_j \): the ordering cycle time of the \( j \)th supplier
- \( x_{ij} \): binary variable equal to 1 if item \( i \) is order from the \( j \)th supplier, otherwise it is 0
- \( k_{ij} \): integer variable that determines the replenishment frequency for item \( i \) in \( j \)th supplier

Same as the classic JRPs, the objective is to minimize the total cost for the downstream buyer. Based on the assumptions mentioned early, the total cost can be given as follows:

\[
\text{Minimize } TC = \sum_{j=1}^{M} \sum_{i=1}^{N} \left( A + \frac{a_{ij}}{k_{ij}} \right) x_{ij} + \frac{1}{2} \sum_{j=1}^{M} \sum_{i=1}^{N} T_j D_i h_i x_{ij} k_{ij}
\]  

The first term of the right hand side represents the costs of ordering various items from the corresponding suppliers

\[
\sum_{j=1}^{M} x_{ij} = 1 \quad i = 1, \ldots, N
\]

Equation (2) constraints each item can only be sourced from one supplier. The replenishment common cycle time for the items supplied by supplier \( j \) can be determined by the following equation.

\[
T_j^* = \left( 2 \left( A + \sum_{i=1}^{N} \frac{a_{ij}}{k_{ij}} x_{ij} \right) / \left( \sum_{i=1}^{N} D_i h_i x_{ij} k_{ij} \right) \right)^{1/2}
\]

4. **THE PROPOSED PSO ALGORITHM**

4.1 **INTRODUCTION OF THE PSO ALGORITHM**

The principles of the PSO algorithm can be summarized as follows:
1. Each particle has its own position which presents a unique solution and moves at a velocity which presents
the change pattern of the solution.
2. Each particle knows its current position and the corresponding objective value at that position.
3. Each particle remembers its own previous position and its corresponding objective value.
4. Each particle is able to generate a neighborhood from any position. The particle has knowledge of the best
position among the swarm and its objective value.

The following equations proposed by Shi and Ebehatr [15] are employed to iteratively modify the velocities of the
particles during searching for a better position (solution):

\[ V_{id} = W \times V_{id} + C_1 \times \text{Rand} \times (P_{id} - X_{id}) + C_2 \times \text{Rand} \times (P_{gd} - X_{id}) \] (4)

\[ X_{id} = X_{id} + V_{id} \] (5)

Where \( V_{id} \) is the velocity of the particle \( i \) in iteration \( d \), \( X_{id} \) is the particle position in iteration \( d \), \( W \) is the inertia
weight, \( C_1 \) and \( C_2 \) are positive constants, called acceleration coefficients. \( \text{Rand} \) are two independently generated
random numbers in the range \([0, 1]\), \( P_{id} \) is the best position the particle \( i \) has reached previously, \( P_{gd} \) is the best
position found among all of the particles. The pseudo code in our paper is as follow:

Begin
\( d = 0 \)
Initialize position and velocity for each particle.
Use the RAND method to Calculate the fitness values for each particle using Eq(1).
Update \( P_{id} \) and \( P_{gd} \) for each particle.
For each \( i \)
Update \( V_{id} \) using Eq(4).
Update \( X_{id} \) using Eq(5).
\( d = d + 1 \)
Update \( P_{id} \) and \( P_{gd} \).
End For
End Repeat

Figure 1: An example of applying particle swarm to a JRP with 10 items sourced from 3 suppliers.

### 4.2 Solution Representation

For the above multi-supplier joint replenishment problem, there are \( N \) items that are sourced from individual \( M \)
suppliers. For each item, one of the \( M \) suppliers is selected to be sourced from. Determining the appropriate suppliers
for all the items involves searching a best combination among MN solutions. To apply PSO to this problem, an $1 \times N$ array can be used to represent that N particles for the N items are simultaneously searching for a best solution. The value at each element of the array (the position each particle can visit) is an integer representing which supplier the corresponding item is sourced from. Fig.1 illustrates an example of ordering ten different items from three individual suppliers. In this example, we set an array with ten places for recording the positions the particles are staying during traveling in the solution space. The value in the array represents the supplier where the corresponding item is sourced from. In Fig.1, the current position of the 10 particles is stored as $(1, 3, 2, 1, 1, 3, 2, 2, 1, 3)$, which represents item 1, 4, 5, and 9 are provided by supplier 1; item 3, 7, and 8 is sourced from supplier 2; item 2, 6, and 10 are sourced from supplier 3.

4.3 Fitness Value Evaluation

Initially, each particle is randomly assigned with an integer number from 1 to 5 and its initial velocity is set with a real number followed a uniform distribution ranging from -0.25 to +0.25. After initialization, each particle has its own initial position and velocity to start with. The common ordering cycle of the items sourced from the same supplier can be determined by using equation (3), and then the total cost with this replenishment cycle can be obtained by using equation (1). This total cost is used as the fitness function of this solution. For each iteration, the particles come up with a new solution by moving to new positions. The fitness value of the new solution can be obtained as mentioned above.

4.4 Calculate $P_{id}$ and $P_{gd}$

In the proposed PSO algorithm, $P_{id}$ represents the best previous position visited by the particle $i$ after $d$ iterative searching procedures, while $P_{gd}$ represents the best position found so far among all the particles. When the position of particles changes, we will update its fitness value, if $P_{gd}^{d+1}$ is less than $P_{id}^{d}$, $P_{gd}^{d+1} \rightarrow P_{id}^{d}$, if $P_{id}^{d+1}$ is more than $P_{id}^{d}$, we keep the original value. The same logic is applied to update $P_{gd}^{d}$.

4.5 Update velocity and position

The velocity and position are two most important components for implementing the PSO algorithm. The particles change its position depending on the velocity which is determined and updated by using equation (4) and (5). When updated, the position of a particle is initially a real number, but later it is set as an integer ranging between 0 and number of suppliers by taking off its fractional portion.

5. A Numerical Example

5.1 Parameter Setting

In this section, we present a numerical example that illustrates the proposed PSO for solving a multi-supplier JRP. In this example, there are 45 different items ($N = 45$) sourced from 5 suppliers ($M = 5$). The inventory manager would like to determine which item is sourced from which supplier, and further implement joint replenishment of the items supplied from the same supplier. Note that the replenishment cycles of the items supplied by one supplier are independent from the replenishment cycle of the items supplied by the other suppliers. The parameter setting for this joint replenishment problem is based on one of the problem sets adopted by Kaspi and Rosenblatt [5]. The demand rate for each item is generated based on a uniform distribution between 100 and 100,000. The major ordering cost for each replenishment is $10. The minor ordering cost for that items purchased from different suppliers have different price was generated from a uniform distribution between $0.5 and $5. The holding cost for each item was also generated from a uniform distribution between $0.2 and $3. For the proposed PSO algorithm, the acceleration coefficients $C_1$ and $C_2$ are set 2 and 1.9, respectively. The inertia weight ($W$), which controls the impact of the previous history of velocities on the current one, is set 0.7. Searching process is terminated after performing 700 iterations.
Table 1: The Results of where items are sourced from and their common replenishment cycles

<table>
<thead>
<tr>
<th>Supplier j</th>
<th>$T_j$</th>
<th>$TC_j$</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.00447</td>
<td>2782.49</td>
<td>14,25,27</td>
</tr>
<tr>
<td>S2</td>
<td>0.00604</td>
<td>1755.39</td>
<td>7,36,42</td>
</tr>
<tr>
<td>S3</td>
<td>0.00444</td>
<td>14475</td>
<td>3,4,5,9,11,12,13,15,18,21,22,23,24,27,31,33,34,39,41,45</td>
</tr>
<tr>
<td>S4</td>
<td>0.00607</td>
<td>6207.31</td>
<td>1,2,8,19,32,35,40,42,</td>
</tr>
<tr>
<td>S5</td>
<td>0.00526</td>
<td>7868.63</td>
<td>6,10,16,17,20,26,29,30,37,38,44</td>
</tr>
<tr>
<td><strong>Total cost</strong></td>
<td></td>
<td>33088.8</td>
<td></td>
</tr>
</tbody>
</table>

5.2 Computational Results

Table 1 presents the results obtained by the proposed PSO algorithm. The items on the most right column of the table are sourced from the corresponding supplier which is list on the most left column. For instance, item 14, 25, and 27 are supplied by supplier 1. As we can see that 3 items are supplied by supplier 1 and 2 each, and 20 items are supplied by supplier 3. The replenishment common cycle for the items sourced from each supplier is also presented in the table. Fig 2 shows that the comparison of the proposed PSO method and the random search approach, by which 30 solutions (same as the proposed PSO method) are randomly generated in each iteration. Both of the search procedures converge after 150 iterations. It is obvious that the proposed PSO method outperformed the random search approach by saving about 15% total costs under the same number of solution searching trials.

Figure 2: Comparison of random search and the PSO method for a JRP with 45 items and 5 suppliers

6. Conclusions

In this paper, a proposed particle swarm optimization method has been used to solve the multi-supplier joint replenishment problem. A numerical example is illustrated for evaluating the performance of the proposed method. The results obtained in the example are highly encouraging since the proposed PSO method is capable of iteratively finding better solutions with small computational requirements. More experiments for larger problems may be
required for confirming such good performance. Moreover, whether the performance of this PSO approach can be improved with different particle velocity update strategy can be explored in future research.

REFERENCES:


Going Green in Supply Chain Management as a Prospective Improvement

Dr. Rania A.M. Shamah
Associate Professor of Business Administration;
High Institute of Cooperative & Managerial Studies

Abstract:
Companies across all sectors recognize that effective supply chain management (SCM) could play a critical role in their future success especially at green manufacturing. Therefore, this paper provides model for how organizations could go green in Supply Chain Management which may enhance organizations market share, as well as, eliminating waste and variation performance and produced products as a prospective improvement. Furthermore, increased attention needs to be paid to assessing how well a GSCM model performs in providing value to stakeholders and the organization as a whole in both quantitative or qualitative measure will be used to address multiple and varied stakeholders needs concerns and to measure organizations success. The applied study is in the Egyptian Air Condition (AC) companies.

Key Words:
Green supply chains, Sharing knowledge, Improvement.

1. Introduction:

Much of the recent supply chain literature has focused on different strategies used to enhance supply chain performance. Strategies such as collaborative planning, forecasting, and replenishment (CPFR), vendor managed inventory (VMI), third-party logistics (3PL), and efficient consumer response (ECR) are just a few of the strategies examined. Research on the environmental and market structure considerations contributing to the complexity of supply chain management and the behaviors of key participants has been somewhat less robust when compared to supply chain strategy research [1]

Nevertheless, organizations competitiveness depends on its ability to perform well in dimensions such as cost; quality; delivery dependability and speed, innovation and flexibility to adopt itself to variation in demand [2]. Therefore, organizations in order to compete well should got: quality beyond the competition; technology before the competition; and costs below the competition [3]; [4]. Hence, continuous improvement is organizations key factor for enhancing its competitiveness adage, while continuous improvement is a companywide process of focused and continuous incremental innovation [5]; [6].

Hence, efficient and effective organization performance is achieved through reasonable use of existent resources. However, it is important for organizations to look at the differences between its competitors to determine the cause of these differences, and propose alternatives to eliminate these differences. The main obstacle in learning has been overcoming resistance to change in order to implement benchmarking. Some organizations do not think they can learn from others. Other issues, which have been identified as barriers, were time constraints, competitive barriers, and lack of personnel resources [5]; [3].

Therefore, between firms, new linkages are being formed to achieve and maintain competitive advantage in the marketplace. These linkages require organisations to move towards network forms and alliances [7]; [8].

In addition, the challenge facing companies today is not just to take up a supply chain management (SCM) initiative but to implement it successfully as the future shall see a war between supply chains [9].

Following an unprecedented growth in global competitiveness over the last few decades, companies all over the world are now taking steps to be ahead of competition in producing world class quality and providing excellent service. In the way, they are also doing this to be a part of a new breed of world-recognized companies called the environmentally responsible companies [10].

Indeed, green supply chain management is transforming the theory and practice of inventory management and the procedures of producing products.

Therefore, this paper argues that the intersection between organization value, sharing knowledge; and environmental performance could enhance supply chains improvement while going green. Although, this paper develops a model for determining key factors affecting green supply chains. The main questions posed in this study are:

1. Is GSCM applicable at Egyptian Ac companies? If so,
2. How Egyptian AC companies acquire their capabilities and resources for going green in supply chains management? If so,
3. What changes needs to be done at AC companies to be able to go green? and Finally,
4. Does GSCM play a critical role in organizations value?

This study has the potential to contribute to a better-informed formulation of economic policies at national and global level. Therefore, the main purpose of this study is to investigate the implications of green supply chain management for the macroeconomic phenomenon of business cycles in Egypt.

Corresponding author: Tel.: +2 011 44 85 85 2; E-mail: rania_shamah@yahoo.com
Finally, is to build a model for enhancing improvement as well as SCM success while, going green by the integration between organization value, sharing knowledge; and environmental performance.

2. Research Hypotheses:

Based on the nature and the purpose of this study, the qualitative method applies to the project work based on the essay format. The other is the quantitative method based on numerical scoring and grading. Finally, the results clubbed together in the mixed approach, a natural choice. In addition, the study is model- interview guide spread over a period of one year submitted to AC companies working in Egypt (Appendix 1). It involved one type of questionnaire, provided across production and purchasing managerial levels AC companies “Top; Senior; and Executive managers”, this questionnaire is divided to four main session: the first session is considered about Organization Value, since the second session is related to Trust atmosphere; while the third session is related to Environmental Performance; and finally the latest session is focusing on Sharing Knowledge. Thus, the questionnaire included questions that overlapped into both qualitative and quantitative approaches. This gave the interviews options to respond qualitative, quantitative, a combination of both or just one of them. Therefore, study hypotheses are:

& “H₁”: Revised model of going green in supply chain management as a prospective improvement is a valid tool to explain in Egypt.
& “H₂”: There is a significance difference between purchasing & production departments within organization value, environmental performance and sharing knowledge.
& “H₃”: There is a significance correlation between organization value ; environmental performance, and sharing knowledge
& “H₄”: Trust is positively related for going green in supply chains.

3. Green Supply Chain Management (GSCM) Model:

Based on literature review of knowledge management; operational management; supply chain management and the organization design, this paper argues that the variables that most affect and are, in turn, influencing the greening the supply chain are those related to the organization strategy. This GSCM model depends on three generalized forms of managing as an intersection between organization value, sharing knowledge; and environmental performance.

These variables are fundamental to identify whether the way an SCM manages its attitudes towered achieving the organization goals as shown in figure (1).

Trust is a key to organizational performance because it enables voluntary cooperation. This form of cooperation becomes increasingly important when command and control styles of management are no longer effective. More and more people work in widely dispersed groups, with increases in the number of people that work at home. In these situations, task performance cannot easily be observed. Besides, the nature of work itself is changing in ways that make command and control approaches to motivation increasingly difficult. Work has become more centred around intellectual labour and a lot of work is done in interdependent teams. Management cannot control everything; it is more and more dependent on willing engagement of employees to work well. Cooperation and trust are important conditions in such a work environment [7].

Trustful relations between organizational members can promote voluntary cooperation and extra-role behaviors. Increasing instances of organizational change have also contributed to the rise of trust on the research agenda. Conditions of change heighten the relevance of trust to organizational performance and to the well-being of organizational members [7]; [11]; [12].

First, Organization Value (OV):

The importance of green culture change has provided the impetus for a series of attempts to conceptualize green organizational culture values [13]. The key values which are commonly positioned are that:

&B Organization should incorporate environmental considerations through the entire organization.
&B Such concerns should be considered through the value chain.
&B Economic goals should be tempered
&B A short term perspective should be substituted with an interrogational time frame.
&B Spirituality, morality, smallness and futurity should be embraced; and
&B The environment should be afforded intrinsic valuation and respect.
However, the interpretation of value will vary depending upon whose perspective it is assessed from. For instance, a manager within a business is looking to increase value to benefit the company’s shareholders, while a customer of the business is looking for value from the organization in perhaps terms of price, quality or social/environmental contribution [14].

Value could be defined as “A term frequently used but infrequently understood and of which numerous interpretations exist” [14]; [15]. Although, values could be defined as “The relatively enduring beliefs about what kinds of behaviors or end-states are preferable to others” [16]. While, Dumond (1996); and Zeithaml, (1988) [17]; [18] clarify value as “It equals customer benefits minus customer sacrifices”. Therefore, values is what a group of people share a set of beliefs about the goals that need to be achieved and the means to be used to attain those goals, there is a basis for organization. In fact, without some common beliefs or values, organizations could not exist; people need a common set of beliefs to come together and create social organizations [16].

However, it will create a value- based management as it is “A paradigm that considers, as a single entity, the organization’s entire chain of activities: those with suppliers, internal functions and customers”. Emphasis is placed on integration of the activities rather than the functions or organizations themselves which explains a value chain as insulted in figure (2) [17]; [19]; [20], the value chain, depicts series of processes which transforms specifications to finished deliverables; although there is a basic sequence of processes which must be performed; finally it describes the relationship between processes – each is dependent on the “supplier” process and the “customer” process.
Therefore, 4PL’s strength and value adding capacity generally lies in their ability to select and co-ordinate a pool of resource from other factions that creates value in excess of that which may have been created had the role been managed internally within the client organization [14].

Although, a co-opetition [21]; [22] business environment, despite the advantages of inter-organizational knowledge sharing, manufactories may impede knowledge sharing, if they regard their partners as potential competitors and try to protect their core knowledge from opportunistic behavior of their partners [23]. This is mainly because sharing key knowledge or confidential information may increase the competitive advantage of their partners, which may be in conflict with their own interests in a highly competitive market.

In addition, the core competence of the organization is to customize knowledge for the postal and distribution market. In practice this means that the organizations raison d’être is dependent on its ability to leverage all elements of the knowledge value chain, including the acquisition, shaping, enhancing, deployment, application, sharing and retention of knowledge on behalf of its clients [24].

Indeed, loyalty and trust within a customer–supplier relationship are needed in order to promote a nurturing partnership, high-level communication, and visibility of information. In certain circles, the option of incentives to encourage co-operation may be considered, the assumption being that a supply chain works well it its companies’ incentives are aligned – that is, if the risks, costs, and rewards of doing business are distributed fairly across the network. There are three reasons why incentive-related issues arise in supply chains [25].

§ When companies cannot observe other firms’ actions they find it hard to persuade these firms to do their best for the network.
§ It is difficult to align interests when one company has information or knowledge that others in the supply chain do not.
§ Incentive schemes are badly designed.

Second, Environmental Performance (EP):

Since the Earth Summit in Rio de Janeiro in 1992, sustainable development has been a philosophy for overall economic development. Those companies that can conserve and manage their resources more efficiently will gain competitive advantage [9]. Under pressure from governments and general public as well as nongovernmental pressures from abroad, an increasing number of enterprises are working to improve their environmental performance [9].

Hence, social and environmental behaviors are sometimes profit-compatible and sometimes not. When parties acknowledge this simple fact, it becomes easier to convince corporations to adopt environmental and social initiatives [26]; [27].

Therefore, to reassure regulatory compliance of their business practices, green manufactures often encourage their supply chain partners to develop an environmental management system (EMS) consistent with the ISO 14000 standards and to obtain the ISO 14001 certification (GEMI- Global Environmental Management Initiative, 2001). The successful implementation of the EMS involves identifying new techniques and opportunities for effective management of environmental impacts [21].

Azzone; and Noci (1998), suggested that the improvement of environmental performance often requires executives to commit significant financial resources in new “cleaner” technologies (financial implications) and to redesign business processes and the corporate organization (managerial and organizational implications). In particular, the effectiveness of environmental programmers’ greatly depends on the executives’ capacity to[28]:

§ Manage environmental innovations jointly with the other competitive factors (time, quality, flexibility, cost, etc.), since the “green” issues are only one of the major challenges faced by companies;
§ Integrate the actions taken by different corporate departments, since environmental programmer have an inter-functional nature, thus requiring the involvement of organizational units characterized by different competencies;
§ Develop new relationships with other firms aimed for solving environmental problems that cannot be managed by a single firm, because of its lack in competencies; and
§ Co-operate with public institutions: experiences carried out in different countries highlight that innovative environmental policies often require their support.

Finally, Sharing Knowledge (SK):

Throughout time, people have learned new skills and accumulated a store of knowledge for them to survive and adapt to challenges. In recent times however the concept of lifelong learning has become embedded in a range
of policies and initiatives. The rise of the “new economy” which is inherently linked to information and communication technology (ICT) has created this impetus for more formalized emphasis on continued or lifelong learning [29].

Nevertheless, the emerging management theories stress the importance of knowledge management within corporate systems, and consider the management of R&D, intellectual property, and technological alliances as key factors to economic growth [30].

Therefore, sharing knowledge is an important component of customer relationship management strategies and that their use can generate significant cost-savings [31]; [32]; [33]. It represents an organized effort to capture expertise and disseminate it to user populations. It also can serve as a support resource for customers (internal or external) who want to solve their own problems rather than rely on technical support staff.

By enhancing learning capacity which refers to the ability of a firm to assimilate and apply new knowledge successfully to its commercial goals [21].

Therefore, GSCM should recognize the main principals of learning [34]: First, Understanding through: mental models or paradigms; Systems thinking, variation, Skills; ability to challenge assumptions; and listen to understand. On the other hand, Processes through: learning cycle: observe, assess, design, and implement (OADI) cycle; and teach others.

4. Testing Hypotheses:

Testing hypotheses \( H \):

Structural Equation Modeling (SEM) using AMOS serves for testing and estimating causal relationships using statistical data as well as qualitative causal assumptions. It is suited for theory testing rather than development of a theory hence is rarely used in exploratory research. It tests the qualitative causal assumptions embedded in the model against the quantitative data in order to confirm the model. One strength of this tool is its ability to estimate latent variables (variables which are not measured directly) from measured variables. SEM allows measuring direct and indirect relationships. [35].

SEM grows out of and serves purposes similar to multiple regression, but in a more powerful way which takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, multiple latent independents each measured by multiple indicators, and one or more latent dependents also each with multiple indicators [36]; [37]; [38]; [39]; [40].

SEM Findings is useful for modeling the complex causal relationship between variables. In this study, SEM was used to examine the model of going green in supply chain management as a prospective improvement in Egypt as figure (3) could explain the interrelation between key factors affecting the Improvement of going green, then followed by figure (4) providing the intersection results.

In this model: Endogenous (Observed) 5 variables measured by average questions: Y, V, Q, Z and IGG, Exogenous (Observed) 7 variables Y2, Y3, Y4,Y5, V2, V3, V4, Q2, Q3, Q4. Exogenous (Unobserved) 5 variables: E1 to E5.

Now estimate the model using AMOS V17.0 applying the data collected. Table (4) shows paths estimates of exogenous and endogenous variables (un-standardized – standardized), standard error, critical ratio and p-value. Table (5) shows covariances and correlations between exogenous variables of the model (values, standard error, critical ratio and p-value).

It is obvious from table (4) that all paths estimated are significance at 1% , then accept paths not equal zero. Although, it is clear from table (5) that estimated paths are significance at 1%, only path between Q2 and Y3 not significance.

Model Fit: Goodness of fit tests: determine if the model being tested should be accepted or rejected. These overall fit tests do not establish that particular paths within the model are significant. If the model is accepted, the researcher will then go on to interpret the path coefficients in the model (“significant” path coefficients in poor fit models are not meaningful).

Additionally, Goodness of Fit should be less than or equal to 1. A value of 1 indicates a perfect fit (Hipp and Bollen, 2003).

Indeed, The values indicated in this model shows a relatively good fit such as: goodness-of-fit index (GFI) = 87.9%, adjusted goodness-of-fit index (AGFI) = 85.5%, comparative-of-fit index (CGFI) = 93.2%, and root mean square error of approximation (RMSEA)=0.067.
Figure (3) Path Model for Explaining the Interrelation between Key Factors Affecting the Improvement of going green

Figure 4: Path Model, Using Structural Equation Modeling
We can add the structural equations model (SEM):

\[ Y = b_1 Y_2 + b_2 Y_3 + b_3 Y_4 + E_1 \]  

\[ V = b_4 V_2 + b_5 V_3 + b_6 Y + E_2 \]  

\[ Q = b_7 Q_2 + b_8 Q_3 + E_3 \]  

\[ Z = b_9 Y + b_{10} Q + E_5 \]  

\[ IGG = b_{11} Y + b_{12} V + b_{13} Z + E_4 \]  

Table 4: Paths Estimates of the Model

<table>
<thead>
<tr>
<th>PATH</th>
<th>B</th>
<th>BETA</th>
<th>S.E.</th>
<th>C.R.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>---</td>
<td>Y2</td>
<td>0.395</td>
<td>0.465</td>
<td>0.013</td>
</tr>
<tr>
<td>Y</td>
<td>---</td>
<td>Y3</td>
<td>0.255</td>
<td>0.339</td>
<td>0.012</td>
</tr>
<tr>
<td>Y</td>
<td>---</td>
<td>Y4</td>
<td>0.276</td>
<td>0.336</td>
<td>0.010</td>
</tr>
<tr>
<td>Y</td>
<td>---</td>
<td>Y5</td>
<td>0.256</td>
<td>0.453</td>
<td>0.011</td>
</tr>
<tr>
<td>Q</td>
<td>---</td>
<td>Q2</td>
<td>0.446</td>
<td>0.502</td>
<td>0.022</td>
</tr>
<tr>
<td>Q</td>
<td>---</td>
<td>Q3</td>
<td>0.431</td>
<td>0.539</td>
<td>0.020</td>
</tr>
<tr>
<td>Q</td>
<td>---</td>
<td>Q4</td>
<td>0.444</td>
<td>0.500</td>
<td>0.021</td>
</tr>
<tr>
<td>V</td>
<td>---</td>
<td>V2</td>
<td>0.385</td>
<td>0.501</td>
<td>0.020</td>
</tr>
<tr>
<td>V</td>
<td>---</td>
<td>V3</td>
<td>0.405</td>
<td>0.562</td>
<td>0.018</td>
</tr>
<tr>
<td>V</td>
<td>---</td>
<td>V4</td>
<td>0.388</td>
<td>0.531</td>
<td>0.019</td>
</tr>
<tr>
<td>Z</td>
<td>---</td>
<td>Q</td>
<td>0.656</td>
<td>0.780</td>
<td>0.032</td>
</tr>
<tr>
<td>Z</td>
<td>---</td>
<td>Y</td>
<td>0.256</td>
<td>0.289</td>
<td>0.034</td>
</tr>
<tr>
<td>V</td>
<td>---</td>
<td>Y</td>
<td>0.119</td>
<td>0.158</td>
<td>0.034</td>
</tr>
<tr>
<td>IGG</td>
<td>---</td>
<td>Z</td>
<td>0.243</td>
<td>0.337</td>
<td>0.011</td>
</tr>
<tr>
<td>IGG</td>
<td>---</td>
<td>V</td>
<td>0.298</td>
<td>0.354</td>
<td>0.013</td>
</tr>
<tr>
<td>IGG</td>
<td>---</td>
<td>Y</td>
<td>0.442</td>
<td>0.694</td>
<td>0.010</td>
</tr>
</tbody>
</table>

\( \chi^2 = 1344.97, \text{ df} = 46, \text{ P} = 0.000, \text{ GFI} = 0.879, \text{ AGFI} = 0.855, \text{ CFI} = 0.932 \text{ and RMSEA} = 0.067. \)

Table 5: Covariances (Cov.) and Correlations (Corr.) of the Model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Y3</td>
<td>Y4</td>
<td>0.670</td>
<td>0.088</td>
<td>7.638</td>
<td>0.000</td>
</tr>
<tr>
<td>Y2</td>
<td>Y4</td>
<td>0.490</td>
<td>0.074</td>
<td>6.597</td>
<td>0.000</td>
</tr>
<tr>
<td>Y2</td>
<td>Y3</td>
<td>0.786</td>
<td>0.090</td>
<td>8.687</td>
<td>0.000</td>
</tr>
<tr>
<td>V2</td>
<td>V3</td>
<td>0.375</td>
<td>0.054</td>
<td>6.947</td>
<td>0.000</td>
</tr>
<tr>
<td>Q2</td>
<td>Q3</td>
<td>0.773</td>
<td>0.089</td>
<td>8.634</td>
<td>0.000</td>
</tr>
<tr>
<td>Y3</td>
<td>Q2</td>
<td>0.012</td>
<td>0.029</td>
<td>0.402</td>
<td>0.688</td>
</tr>
<tr>
<td>Y4</td>
<td>V3</td>
<td>0.067</td>
<td>0.039</td>
<td>1.720</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Squared multiple correlations for structural equations (R2): is the portion of output which gives the percent of the variance in the latent dependent variable(s), accounted for by the latent independent variables. Table 6 shows R-squared for all endogenous variables, the explained variance of Q, Y, Z, V, IGG are 94.4%, 97.9%, 69.2%, 90.3% and 95.3% respectively.
Table 6: R Squared for all Endogenous Variables

<table>
<thead>
<tr>
<th>Var.</th>
<th>Value</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>0.944</td>
<td>3</td>
</tr>
<tr>
<td>Y</td>
<td>0.979</td>
<td>1</td>
</tr>
<tr>
<td>Z</td>
<td>0.692</td>
<td>5</td>
</tr>
<tr>
<td>V</td>
<td>0.903</td>
<td>4</td>
</tr>
<tr>
<td>IGG</td>
<td>0.953</td>
<td>2</td>
</tr>
</tbody>
</table>

Therefore, from previous results: "$H_1\$" is accepted, "Revised model of going green in supply chain management as a prospective"

Testing hypotheses "$H_2\$":

Multivariate Analysis of Variance (MANOVA) Test:

Multivariate analysis of variance (MANOVA) is used to test this hypothesis. MANOVA technique is an extension of common analysis of variance (ANOVA). In ANOVA, differences among various group means on a single-response variable are studied. In MANOVA, the number of response variables is increased to two or more. The hypothesis concerns a comparison of vectors of group means. The multivariate extension of the F-test is not completely direct [41]; [42]. We fitting a general linear statistical model relating 3 dependent variables (organization value, environmental performance and sharing knowledge) to 1 predictive factors of depranment type (purchasing – production).

The analysis shows several multivariate ANOVA (MANOVA) statistics for each of the effects in the model. These statistics are used to determine whether a particular effect has a significant relationship with the group of dependent variables being modeled. P-values below 0.05 indicate that an effect is statistically significant at the 95.0% confidence level. In this case, were found no significant difference according to Pillai's Trace test (reject "$H_2\$"). The partial Eta Squared values in table 10 represent the most small weknees important facotr (less than 1%).

Table 7: MANOVA Test Results

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai's Trace</th>
<th>F</th>
<th>Hyp. df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.923</td>
<td>823.025</td>
<td>3</td>
<td>207</td>
<td>0.000</td>
<td>0.923</td>
</tr>
<tr>
<td>Dep. type</td>
<td>0.007</td>
<td>0.452</td>
<td>3</td>
<td>207</td>
<td>0.716</td>
<td>0.007</td>
</tr>
</tbody>
</table>

It is obvious from previous table that there is no significant effect at 0.716 ; and Partial Eta² 1% related to department type “Production; and Purchasing” on organization value; environmental performance; and knowledge share

It is obvious from table (8) that there is no significant effect related to department type “Production; and Purchasing” on organization value; environmental performance; and knowledge share, therefore "$H_1\$" is rejected.
Table 8: Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Dep. Var.</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>OV</td>
<td>0.921</td>
<td>1</td>
<td>0.921</td>
<td>1.343</td>
<td>0.248</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>0.242</td>
<td>1</td>
<td>0.242</td>
<td>0.518</td>
<td>0.472</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>0.520</td>
<td>1</td>
<td>0.520</td>
<td>0.697</td>
<td>0.405</td>
<td>0.003</td>
</tr>
<tr>
<td>Intercept</td>
<td>OV</td>
<td>1184.735</td>
<td>1</td>
<td>1184.735</td>
<td>1727.706</td>
<td>0.000</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>1105.965</td>
<td>1</td>
<td>1105.965</td>
<td>2369.548</td>
<td>0.000</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>944.235</td>
<td>1</td>
<td>944.235</td>
<td>1265.962</td>
<td>0.000</td>
<td>0.858</td>
</tr>
<tr>
<td>Dep. type</td>
<td>OV</td>
<td>0.921</td>
<td>1</td>
<td>0.921</td>
<td>1.343</td>
<td>0.248</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>0.242</td>
<td>1</td>
<td>0.242</td>
<td>0.518</td>
<td>0.472</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>0.520</td>
<td>1</td>
<td>0.520</td>
<td>0.697</td>
<td>0.405</td>
<td>0.003</td>
</tr>
<tr>
<td>Error</td>
<td>OV</td>
<td>143.317</td>
<td>209</td>
<td>0.686</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>97.549</td>
<td>209</td>
<td>0.467</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>155.886</td>
<td>209</td>
<td>0.746</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>OV</td>
<td>2267.278</td>
<td>211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>2046.715</td>
<td>211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>1838.952</td>
<td>211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>OV</td>
<td>144.238</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>97.791</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SK</td>
<td>156.405</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Testing hypotheses "H₃":**
Pearson Coefficient Correlation is used for testing this hypothesis.

Table 9: Pearson Correlation between organization value; environmental performance; and sharing knowledge

<table>
<thead>
<tr>
<th></th>
<th>OV</th>
<th>EP</th>
<th>SK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. Var.</td>
<td></td>
<td>0.691**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.795**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

From tables (9) it is obvious that there is a significant increasing proportional relation between organization value; environmental performance; and sharing knowledge for going green in supply chains; Therefore "H₃" is accepted.

**Testing hypotheses "H₄":**
Pearson Coefficient Correlation is used for testing this hypothesis.

Table 10: Pearson Correlation between Trust and Going Green

<table>
<thead>
<tr>
<th></th>
<th>Going Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td>0.863**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

From tables (10) it is obvious that there is a significant increasing proportional relation between trust and going green in supply chains; Therefore "H₄" is accepted.

5. Results and Recommendations:
Based on this study, it is considered that the key factor for successful improvement in AC organizations for going green in supply chain management are generating new organization value, sharing knowledge; and environmental performance. To examine the hypothesis a field study technique was employed. The following are the main results of this study:
AC organizations gain an opportunity for going green. Which, reflect a supported organization culture, for integrating green manufacturing within organizations’ strategy, goals; and products.

There are, and still will be employees who do not have a solid grasp on their duties when going green. Therefore, seminars; value aids should be used to provide leaders experience when applying green manufacturing and focusing on their benefits to stakeholders “ employees, customers; suppliers; and shareholders”; and

AC organizations leaders; and top managers must focus on going green as a social requirement, for avoiding employee resistance. Hence, greening a supply chain is as an enterprise-wide goal.

6. Conclusion:

The purpose of “Green Supply Chain Management Model” is to gain a better understanding of how could applying organization value and to share knowledge between partners for enhancing quality and lead to an improvement which will be continuously by adding value to the stakeholders as well as to the community. The “Green Supply Chain Management Model” has been proposed to analyze the factors affecting improvement at manufacturing organizations.

Wherein, today’s competencies become tomorrow’s core rigidities with unprecedented speed. An organization should have the capacity to exploit its resources and learning capabilities better than its competitors, if it decides to assume a given competitive strategy.

To conclude Green Supply Chain Management is a process that helps organizations find, select, organize, disseminate, and control its resources to gain business advantage through environmental phenomena.

However, the strength of this study methodology lies in its comprehensive coverage of various aspects of Green Supply Chain Management and its implementation at AC manufactories. It provides for both, as in-positivist researchers adopt a quantitative methodology and carry out surveys and questionnaires. Furthermore, interpretive researchers adopt a qualitative methodology and carry out interviews and ethnographies.

On the other hand this is Limitations; the study period interval in data collection may have influenced the variance in responses and therefore should be considered a limitation.

In addition, Due to many incomplete responses that were received and the qualitative response parts are sometimes estimated based on collected impressions, there is a minor influence on the accuracy of the estimates for “key areas of weakness” in green implementation. While these limitations outline potential areas of weakness in the methodology, yet, it still has been possible to undertake a comprehensive approach successfully.

Greening the supply chain survey identified a significant degree of impact on the awareness of the average employee regarding EMS. The need for formulating an overall strategy for knowledge base to support EMS comes forward very strongly.

The following factors are important for the future requirements to ensure going green in supply chains initiatives to succeed:

- High priority top management support;
- Establishing unique organizations value to support applying EMS in SCM;
- Developing and coordinating well communications plans; and
- Establishment of incentives to going green.

7. References:


Appendix 1: Population and Sample Size

Air Condition companies working in Egypt are around twelve companies laid in Private sector. A Pilot study of 30 units is used for evaluating variables validity and reliability. Inter-consistency is used for examining variables validity; although; Alpha-Cheloficiof is used for examining variables reliability. From those tests results are as follows:

**AC Sample:**
All variables are high validated, as Alpha-Cheloficiof = 0.963

**First; Survey Society:** As mentioned before, one questionnaires was designed this study, submitted to those who work in AC “Purchasing and Production” Department.

**Second; Sampling Society:** Stratum Random Sample is carried as follows:
The following table provides AC companies names working in Egypt.

<table>
<thead>
<tr>
<th>No.</th>
<th>Company Name</th>
<th>No.</th>
<th>Company Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Al- Gazira for Ac and Refrigerators</td>
<td>2</td>
<td>Union Air</td>
</tr>
<tr>
<td>3</td>
<td>Al – Sabah</td>
<td>4</td>
<td>Egypt Company for AC Miraco</td>
</tr>
<tr>
<td>5</td>
<td>Hamam Manufactory</td>
<td>6</td>
<td>Al- Sawlagan Manufactory</td>
</tr>
<tr>
<td>7</td>
<td>National Company for AC</td>
<td>8</td>
<td>Marly for AC</td>
</tr>
<tr>
<td>9</td>
<td>SRC Company for AC</td>
<td>10</td>
<td>Marlow Manufacturing</td>
</tr>
<tr>
<td>11</td>
<td>Mobil Clement Control Company</td>
<td>12</td>
<td>Bahgat Group</td>
</tr>
</tbody>
</table>

**Third, Determining Study Sample through:**
For calculating sample size: (Seheaefer; Mendenhall &Lyman, 2000) suggested an equation is used as followed:

\[
 n = \frac{Z^2PQ}{D^2} = \frac{1.96^2(0.5)(0.5)}{0.07^2} = \frac{0.9604}{0.005} \cong 192 \text{ at least}
\]

Where:
- P : proportion in population, when unknown = 0.5, Q=1-P,
- D : deviation between parameter and statistics.
- Response Rate=211/220=95.9%

**Descriptive Data**
Number of demographic questions was asked of respondents in the introduction session of the questionnaire.

1. Department
The following table and figure provides a breakdown of respondents’ department representation of the sample.

**Figure 1 Department Distribution**

<table>
<thead>
<tr>
<th>Dep.</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pur.</td>
<td>37</td>
<td>17.5</td>
</tr>
<tr>
<td>Pro.</td>
<td>174</td>
<td>82.5</td>
</tr>
<tr>
<td>Total</td>
<td>211</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The purchasing to production ratio in the sample is 17.5% to 82.5% respectively.
2. Experience

The following table provides a breakdown of respondents’ experience representation of the sample. The average (Mean) years of experience is 11.7 with standard deviation (S.D.) 3.9, minimum and maximum are 1, 19 respectively.

Table 3: Descriptive Statistics of Years of Experience

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>11.67</td>
<td>3.85</td>
</tr>
</tbody>
</table>

Reliability Test of the Questionnaire's Items

To begin the analysis, Cronbach’s alpha method is used to check the reliability of the attitude dimensions of the questioner. Ideally, the alpha coefficient of a scale should be above 0.7, though this coefficient could be lower depending on the number of factors and length of the scale used. Table 3 shows all the alpha values of the attitude dimension is 0.963, the alpha coefficient varies from a low of 0.740 to a high of 0.935 these results are to a certain extent reliable.

Table 4: Cronbach's alphas of the Questionnaire's Items:

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cronbach Alpha</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization Value (Y)</td>
<td>0.935</td>
<td>21</td>
</tr>
<tr>
<td>Definitions (Y1)</td>
<td>0.890</td>
<td>11</td>
</tr>
<tr>
<td>Shared Value (Y2)</td>
<td>0.828</td>
<td>4</td>
</tr>
<tr>
<td>Communication (Y3)</td>
<td>0.828</td>
<td>3</td>
</tr>
<tr>
<td>Power (Y4)</td>
<td>0.689</td>
<td>3</td>
</tr>
<tr>
<td>4PI (Y5)</td>
<td>0.666</td>
<td>4</td>
</tr>
<tr>
<td>Trust (T)</td>
<td>0.817</td>
<td>7</td>
</tr>
<tr>
<td>Environmental Performance (V)</td>
<td>0.821</td>
<td>11</td>
</tr>
<tr>
<td>Definitions (V1)</td>
<td>0.784</td>
<td>5</td>
</tr>
<tr>
<td>Opportunistic behavior (V2)</td>
<td>0.744</td>
<td>3</td>
</tr>
<tr>
<td>Social Performance (V3)</td>
<td>0.728</td>
<td>3</td>
</tr>
<tr>
<td>Resource Fitness (V4)</td>
<td>0.722</td>
<td>4</td>
</tr>
<tr>
<td>Sharing Knowledge (Q)</td>
<td>0.895</td>
<td>12</td>
</tr>
<tr>
<td>Definitions (Q1)</td>
<td>0.834</td>
<td>6</td>
</tr>
<tr>
<td>Participation (Q2)</td>
<td>0.740</td>
<td>3</td>
</tr>
<tr>
<td>Knowledge Base (Q3)</td>
<td>0.742</td>
<td>3</td>
</tr>
<tr>
<td>Learning Capacity (Q4)</td>
<td>0.723</td>
<td>4</td>
</tr>
<tr>
<td>Overall</td>
<td>0.963</td>
<td>51</td>
</tr>
</tbody>
</table>
Tool support for industrial engineering in small enterprises

Egon Müller, Sebastian Horbach*, Jörg Ackermann

Department of Factory Planning and Factory Management
Institute of Industrial Sciences and Factory Systems
Chemnitz University of Technology
Chemnitz, Saxony, D-09126, Germany

ABSTRACT

Competence Cells represent the smallest reasonable autonomous unit of production. They work together in non-hierarchical regional networks. Thus, Competence-Cell based networks constitute a special form of enterprise networks. The planning of logistics structures and production plants in networks of small and medium sized enterprises depends on an effective tool support. The Net Planning Assistant was developed to give such support to Competence Cells working in competence networks. However, it is also suitable for traditional enterprises and their networking. In order to meet its requirements, the Net Planning Assistant has a modular structure. The concept of the Net Planning Assistant embraces three main components: the production database, the associated software tools and the uniform interface concept. The production database is the instantiation of the production data model. The production data model was developed from a view concept with the main views structural element, targets, yield, functions and resources/competences. For the administration of the production database a web interface has been developed. The Net Planning Assistant comprises a wide range of commercial and proprietary software tools, associated to the production database and to each other. Different types of tools such as small stand-alone applications, internet based solutions, or web services can be incorporated. The software tools cover different fields, which are relevant to the planning of logistics structures and production plants such as modelling, optimisation, simulation, or participative planning. A uniform interface concept was developed to connect the software tools to the production database. There are three basic ways of transferring the data between the production database and the tool. The exchange of data between the different tools is realised via the production database. Several examples for associated software tools and their interaction on the base of the uniform interface concept are given and areas of their deployment are shown.

1. INTRODUCTION

Autonomous, elementary units of production, co-operating in temporary networks, are viewed as a key organisational form of enterprise in the 21st century. A scientific approach is provided by non-hierarchical networks based on customer-oriented, directly linked, smallest autonomous service units called Competence Cells. Simultaneously, this concept, which is researched at Chemnitz University of Technology, points out perspectives for present-day small and medium-sized enterprises (SME) to face ever-changing economic conditions. [1]

Competence-Cell-based networks constitute a special form of enterprise networks. Special requirements on the planning of logistics structures and production plants in such networks result from the specifics of the approach. To meet those requirements, a Planning Concept for Networks (PlaNet) has been developed [2, 3]. PlaNet has a modular structure. The component of PlaNet which represents its implementation as a software concept and offers the necessary tool support is the Net Planning Assistant (NPA).

The requirements on the planning in Competence-Cell-based networks widely apply also to current networks of SME. Therefore, the application of PlaNet including NPA is expected to be beneficial to existing SMEs.

First, the general concept of the Net Planning Assistant is discussed. Then, the components of NPA are introduced in more detail. Finally, the state of its development is presented.

* Corresponding author: Tel.: +49-371-531 35184; Fax.: +49-371-531 835184; E-mail: sebastian.horbach@mb.tu-chemnitz.de
2. Net Planning Assistant

For an effective planning of logistics structures and production plants in Competence-Cell-based networks, adequate software support is vital. There are a number of special requirements on the suitable software environment in such networks. Such requirements are a holistic treatment, easy usability, interconnection of the Competence Cells, support of participative planning, extensibility and low costs.

There are two possible approaches for a software environment for such a complex systems as Competence-Cell-based networks. On the one hand, an expensive monolithic system can be installed [4, 5]. On the other hand, a number of smaller tools, each of them specialising on a couple of tasks, can be integrated [6].

The advantage of the monolithic system is the better integration of its modules and the existence of a uniform user interface. Disadvantageous are especially the high costs of such a system not only for its actual purchase, but also for the installation and the training of the users. Usually, those costs arise all at a certain point of time. In many cases also a difference in the quality of the different modules can be observed.

The collection of tools however can be assembled exactly to the needs of the SME. Since they can be purchased one at a time, the costs at one period can be kept low, and the solution can be gradually extended. The best tool in each field can be picked, so that for each area the best expertise is used. The disadvantage of such a collection lies especially in the effort, which is necessary for the integration of tools.

For the tool support in planning of logistics structures and production plants in Competence-Cell-based networks, the second approach was used. A concept for a Net Planning Assistant (NPA) was developed on this base. It is illustrated in Figure 1. The Net Planning Assistant builds on Planungssystem 2000+, which is a concept for establishing a software pool for a logistic and simulation centre with project specific configurations. This concept was adapted to the needs of networks of SME. On the one hand, tools for additional fields of applications had to be incorporated. On the other hand, the development of software technology, especially web programming, had to be taken into consideration.

![Figure 1: Concept of the Net Planning Assistant](image-url)
In the centre of the NPA is the production database (PDB). Through a uniform interface concept, the PDB associates commercial and self-developed software for the corporate realisation of complex planning processes.

Every Competence Cell uses exactly the tools and interfaces which it needs. Deployment strategies have been already presented in [7].

3. COMPONENTS OF THE NET PLANNING ASSISTANT

3.1 PRODUCTION DATABASE

The Production Database (PDB) is the central data storage place of NPA. The PDB is the implementation of the production data model (PDM). The PDM was modelled according to a view concept, which was introduced by Näser, Ackermann and Baum [8] (see also [9]).

The view concept (Figure 2) is a framework for the mapping of socio-technical systems. It contains the views

- Target view,
- Yield view,
- Function view,
- Competence and Resource view,
- Structure view.

Figure 2: View Concept [8]
The target view comprises the objectives of the system. The necessary yield to reach those targets is mapped in the yield view. Yield can be products as well as services. The production of the yield is realized by functions. To exercise those functions resources and the competences of the workforce are necessary. The relations between the different objects are mapped in the structure view. ([8])

The view concept can be applied to all levels of a socio-technical system, hence, also to all levels of a production system.

In the production data model first the views were mapped. They had to be complemented by the object type structure unit which represents the different types of socio-technical systems, which are described, such as Competence Cell, Competence Network, production system etc. The relations between the views were modelled as entities too, since they have a cardinality of m-n. From those basic objects, further specialisations were derived. Hereby, especially the data structures of the associated software tools were taken into consideration if possible.

The production data model was realised with a case tool as a conceptual data model. Due to the fact that the relational model is still the dominating approach for the majority of industrially applied database management systems (DBMS) it was decided, to use this approach and not an object oriented database model. From that model, databases for different types of DBMS can be generated. Alternatively an XML Schema can be produced.

The elicitation and the processing of the data are mainly done by the associated software. PDB provides the input information to the planning tools and stores their output information. This output can again be used as input for planning tools, which realise succeeding processes. In some cases a direct access on the production data is necessary. To make this more comfortable, a prototypical web-based user interface is under creation. It enables the Competence Cells to access the PDB with their internet browser, thus avoiding the installation of special software.

The uniform interface (Figure 3) concept enables the associated tools to communicate through the production database.

![Uniform Interface Concept](image)

Figure 3: Uniform Interface Concept
3.2. **Uniform Interface Concept**

The interaction of the different software tools is organised through a uniform interface concept. Thereby, the exchange of data can take place in three different ways (Figure 3).

1. **Direct access to the PDB without transformation.** The data access is realised by ActiveX Data Objects (ADO) or if this is not possible by the Open Database Connectivity (ODBC) standard.

2. **Transformation through intermediate files or tables.** An emerging standard for this is the eXtensible Markup Language (XML). Many applications use Microsoft Excel files (.xls) for the exchange of data. A similar approach is to use intermediate tables for transferring data between two databases.

3. **Transformation without intermediate files or tables.** This usually is realised through access of the database, which is used by the application, provided the data structure is disclosed.

The way which has to be chosen depends on the associated software. Methods 2 and 3 require the use of adapters, which realise the transformation. The generation of the adapters should be supported by proper software, which helps to automate the process. The generation of interfaces requires expert knowledge of the data structure.

The uniform interface concept should be particularly applied to the elicitation of the data, which is often already stored electronically.

3.3. **Connected Software Tools**

A number of commercial and proprietary software tools are connected through the PDB in order to solve the existing planning problems. Typical assignments of connected software together with examples of particular solutions are shown in Figure 1. They realise the different functions of the planning process. Some of the software solutions realise components of PlaNet. The empty box in Figure 1 suggests that NPA can be extended to yet unconsidered or unknown fields of application. Special emphasis is placed on tools which support participative planning.

The incorporation of a software solution into the NPA usually requires the generation of two interfaces in accordance with the uniform interface concept.

There are different types of applications which are especially qualified for the NPA. In the past (windows) applications, which have to be installed on every client, were used. Alternatively scripting environments, which are offered by a wide number of applications (the most popular being Visual Basic for Applications) can be deployed to integrate functions in commonly used software solutions.

Preferable are internet-based applications, which support the interlinked planning and can be run through a web browser, ideally on all kinds of systems. Applications which don’t demand any installation on the client’s computer are programmed in languages like PHP or Python. The same is true, when Microsoft’s Active Server Pages (ASP) can be used on the web server.

While these techniques are good for rather simple solutions, applications which use a lot of graphics should be better run on the client computer. First choice for the development of such applications is Java which needs the Java Virtual Machine on the client but is independent from the operating system.

A promising emerging technology is the use of web services. That means that interfaces to functions are provided. Functions, which are implemented by a web service, can be easily called by most programming environments.

4. **Implementation**

Several steps have been taken for the realisation of the concept of the Net Planning Assistant. The evaluation of the NPA is executed in two stages. Firstly, the associated software tools need to undergo a check whether they are suitable for Competence-Cell-based networks. Secondly, it remains to be verified whether the overall concept is fit to meet the requirements. For that purpose, data was collected from a collaboration of SME.

From the conceptual data model, a database in Microsoft SQL-Server was generated, in which an exemplary network of Competence Cells was mapped.
A tool was developed for the implementation of the uniform interface concept, especially for the case that a transformation is necessary (see section 3.2, methods 2. and 3.). This tool named Data Bridge is freeware and can be obtained from IBBSOFT ([10]). An interface to the tool CATRIN ([11]), which solves vehicle routing problems, has been developed.

A number of commercial tools have been used for solving planning problems in Competence-Cell-based networks. Besides the aforementioned CATRIN ([11]), especially 4flow Vista ([12]) and ORionPI ([13]) as tools for the optimisation of logistics structures, and the simulation software Tecnomatix Plant Simulation ([5]) were deployed.

The data management is supported by a number of web applications. Firstly, there is a tool for manual editing, which is suitable if a small amount of data needs correction. Secondly, an instrument for searching Competence Cells by the views of the view concept is available. These tools support also the identification of missing competences in a network. A third application analyses the database for possible errors, and is especially recommended after the collection of raw data. The fourth software is designed for the documentation of the data with the goal of supporting both the analysis through the planner, as well as the presentation to decision makers.

The implementation of those tools uses PHP as the programming language. PHP is easy to learn, widely used, and supported by the vast majority of web servers. The four applications were united to a data management toolset (Figure 4).

![Figure 4: Screenshot of the internet based data management toolset](image)

A number of tools on the basis of MS Excel have been developed. These tools can be useful for solving particular planning tasks. They assist in executing an ABC-Analysis, the determination of the necessary area for a work place, the optimisation of the production program through linear programming, and the optimisation of the structure of a workshop. The latter tool implements a number of algorithms and produces a line structure or a spatial
structure. Since those tools are proprietary, the production database can be accessed without transformation. The necessary interfaces were programmed and included in MS Excel.

The visualisation and participative planning of the productions system is supported by a combined hardware and software solution named visTable ([7, 14]). This tool also supports the direct access to the production database.

5. SUMMARY

With the Net Planning Assistant, proper tool support for the planning of logistics structures and production plants in Competence-Cell-based networks and hence also existing networks of SME is provided. To meet the financial restrictions of the typical participants, a modular structure was chosen. The data needed in the planning process, is stored in a production database. Commercial and proprietary software tools are connected through this database and a uniform interface concept.

The production database was generated and a tool to realize the uniform interface concept was developed. Thus, a number of software tools could be already integrated. In the future more tools need to be incorporated, especially internet based solutions using innovative technologies such as web services.

ACKNOWLEDGMENTS

The work which is presented in this paper has been supported by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG).

REFERENCES

Radio Frequency Identification (RFID) in Manufacturing Enterprises: From Data to Information

Can Saygin* and Jagannathan Sarangapani

1 Department of Mechanical Engineering
Manufacturing Systems and Automation Lab
The University of Texas at San Antonio
San Antonio, Texas 78249
U.S.A.

2 Department of Electrical and Computer Engineering
Embedded Systems and Networking Lab
Missouri University of Science and Technology
Rolla, Missouri 65409
U.S.A.

ABSTRACT

A growing number of organizations are investigating the use of radio frequency identification (RFID) as a tool to improve their business processes across the enterprise. When implemented properly, the benefits RFID brings to supply-chain management, logistics, and asset tracking are clearly understood. Yet many companies do not realize the value RFID can supply by providing data input into a company’s Six Sigma and Lean applications by facilitating operational visibility. This paper presents a case study on deployment of RFID technologies in industry with a focus on the supply chain. A systematic approach that integrates the phases of business process reengineering, RFID data-based decision making, laboratory-level prototyping, and pilot implementation at industry site are discussed. Applications, ranging from inventory management to shipping/receiving operations to emulation of supply chain operations, developed at the Manufacturing Systems and Automation (MSA) laboratory are presented. The study highlights how the lean systems thinking and how RFID technologies can be used as process improvement enablers.

1. INTRODUCTION

Radio Frequency Identification (RFID) has received a great deal of attention for its potential ability to perform non-contact object identification and to provide visibility at the point of use in a variety of different industries [1]. Although RFID is not a new technology as it dates back to the techniques developed to differentiate “friendly” aircraft from enemy warplanes in World War II. However, recent developments in computer technology and electronics have combined to make the RFID technology potentially viable for commercial purposes. As shown in Figure 1, a typical RFID system consists of three components: 1. An electronic data carrying device, called a transponder or tag, 2. Antennas and readers that facilitate tag interrogation, and 3. Software, called middleware, that controls the RFID equipment, manages the RFID data, and distributes information to other remote data processing systems by interfacing with enterprise applications. An RFID system can be considered a wireless communication system since the reader communicates with the tags by using electromagnetic waves at radio frequencies [2]. RFID systems can be categorized as active and passive systems. In an active system, the tag (i.e., active RFID tag) has its own power source, which is a battery, enclosed in the transponder housing. In a passive system, the tag does not have its own power source; instead it draws power from the reader’s radio signals. Passive tags are inexpensive compared to active tags.

Information is the fuel that drives the economy and the society today [3]. As manufacturing operations go increasingly global, proper coordination among business and manufacturing units can be provided by sharing information in a timely manner [1]. Similarly, market and other uncertainties can be reduced and better managed by sharing information instead of building up inventories [4]. From supply chain level operations to shop floor level manufacturing execution, deploying RFID technologies can help facilitate information sharing and provide visibility in processes [5]. Further, with the existence of proper infrastructure, RFID can improve real-time exchange of data between locations and entities in a logistics network, facilitating better and more accurate information flow.

* Tel.: 1 – 210 – 458 – 7614 ; Fax: 1 – 210 – 458 – 6504 ; E-mail : can.saygin@utsa.edu
2. PROBLEM DEFINITION

While the physics of RFID is relatively simple, there are two fundamental problems that must be addressed prior to widespread adoption [6-9]: First, to make readers and tags effectively and efficiently communicate in order to achieve the primary goal of seamless data flow, and second, to re-engineer business processes with RFID data-driven decision making models. Various business cases that involve RFID implementations can be found in the literature.

In this paper, the second issue is addressed through a case study, which involves two levels of deployment:

1. **Facility Level**: The problem at facility level is to develop an RFID-based inventory management methodology for time-sensitive materials in a manufacturing environment.

2. **Supply Chain Level**: The problem at supply chain level is to utilize RFID data for ordering and tracking of time-sensitive materials among the material supplier, materials distributor, and the manufacturing site that uses the materials.

Inventory management of time-sensitive materials is very critical due to their shelf lives. One primary concern is to ensure that the required materials are available at all times to the operators. At the facility level, the lack of time-sensitive materials results in loss of production and, in turn, loss of profits. On the other hand, those materials that are not used in production within their lives expire and become another cost factor. In addition, disposal of time-sensitive materials, once they reach their shelf life, to prevent their usage on a product is also major concern. The problems are reflected at the supply chain level, as well. First, excessive “just in case” inventory is ordered due to lack of data on actual inventory status. Second, business processes among the three business partners in the supply chain depend on paper-based tracking and “too many hands” touching the inventory, which lead too long lead times.

2.1. **Facility Level Inventory Tracking**

The manufacturing environment presented in this case study involves tracking of approximately 5000 time-sensitive inventory items, which involve over 150 varieties, stored in approximately 100 buffers (i.e., each buffer holds 500 inventory items) that are used for assembly of a high-value product. The current practice is to order a higher quantity of materials than necessary determined by the baseline inventory of each buffer in order to attain a high service level, which is defined as the percentage of shop floor orders met on time. Inventory in each buffer is tracked manually. However, due to the pressures to complete the orders on time, operators who do not have the right inventory item in their designated buffers, “borrow” inventory from other buffers. This undesired borrowing of items lead to discrepancy between the manually tracked inventory data and the actual inventory in each buffer, which leads to a lower service level and higher amount of expired materials. Overall, lack of real-time visibility on inventory levels leads to wasteful activities that add cost and increase lead time.

The proposed model to overcome the aforementioned deficiencies involves implementing a facility-wide decision making model that utilizes RFID data coming from tagged inventory items stored in the buffers. The location of each buffer combined with time-stamped tag ID can be used to capture the actual status of each buffer and incoming and outgoing materials, thus inventory replenishment can be carried out in a more efficient manner. The decision making model uses a trend-adjusted exponential smoothing algorithm [1]. It uses two smoothing parameters, $0 < \alpha < 1$ and $0$
Radio Frequency Identification (RFID) in Manufacturing Enterprises: From Data to Information

< β < 1, as coefficients for the average production demand and its trend, respectively. The adaptive inventory scheme looks at the difference between the current inventory level of a particular inventory item at a buffer and the associated forecast (i.e., predicted demand) in order to determine the amount of material that needs to be ordered. Total amount of materials needed is calculated by simply adding the required amount at each buffer. The purchase order is generated automatically due to the availability of RFID-based data and is shared in real-time with the materials distributor. The “operational visibility” gives the materials distributor enough time to plan and replenish buffers more effectively.

In order to demonstrate the validity of the proposed model, a simulation study is carried out, which benchmarks the current practice against the new model. The simulation model is built around a simplified version of the actual manufacturing environment. It includes 18 buffers, 23 different types of materials, a total inventory level of approximately 5000 items, and 10 replications of a production period of 7 months. The simulation study uses six performance measures to evaluate the overall performance of the two inventory models. These are defined as expired, normal use, ordered, stock-out, reallocated, and service level. Expired refers to the total number of time-sensitive materials that expire. Normal Use is the total number of time-sensitive materials that are available in designated buffers and are used in production. Ordered is the total number of time-sensitive materials ordered. Stock-out is the total number of time-sensitive materials requested by an operator but were not in the buffer at the time of the demand, and Reallocated is the total number of time-sensitive materials an operator obtains from a neighboring buffer.

The proposed model is tested via simulation using different values for the two smoothing parameters α and β in order to determine their most effective values. As shown in Table 1, each level-combination of the smoothing parameters, α and β (i.e., sub-models), yields a different performance for different performance measures. For instance, the sub-model, where α=0.8 and β=0.2, yields the best result for Expired. On the other hand, the same sub-model ranks number 8 for Normal Use, Stock-out, and Service-level metrics. Similar comparisons can be made for the other sub-models. The relationship among the performance measures needs to be taken into consideration when analyzing the results. For instance, the more that is ordered the more inventory items are likely to expire and the less operators will run out of stock because more inventory items are potentially available for production. Overall, since α=0.2 and β=0.2 yield the highest service level, which is the primary performance metric identified by the team at the manufacturing site, with the lowest number of reallocations, α=0.2 and β=0.2 combination is selected to be the best combination of all nine sub-models.

Table 1: Smoothing parameters and performance measures

<table>
<thead>
<tr>
<th>α</th>
<th>β</th>
<th>Expired</th>
<th>Normal Use</th>
<th>Ordered</th>
<th>Stock-out</th>
<th>Service Level</th>
<th>Reallocated Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>8*</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1 (96%)</td>
<td>7185</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>14508</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>16700</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>15044</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>9 (91.5%)</td>
<td>18175</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>19257</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>18077</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>19903</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>18930</td>
</tr>
</tbody>
</table>

Difference** 150 2900 2100 2850 4.5% 13000

(*) The lower the relative weight the better (i.e., a lower relative weight is desired)

(**) Difference (i.e., variation of results) between the best (1) and the worst (9) results under each performance measure

After the most effective settings for the smoothing parameters are identified, the proposed model using these settings is benchmarked against the current practice. As shown in Table 2, the proposed model outperforms the current practice with higher service level, fewer amount of expired materials, more items withdrawn from designated buffers (i.e., Normal use), fewer stock outs, and much less ordered. Although there are relatively more reallocated items in the proposed model, it is due to the visibility provided by RFID that facilitates “controlled” inventory sharing; items unavailable at designated buffers are tracked easily and reallocated from other buffers in order to maximize service level.

Page 1521
Table 2: Proposed model ($\alpha=0.2, \beta=0.2$) against current practice

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Service Level</th>
<th>Expired</th>
<th>Normal Use</th>
<th>Stockout</th>
<th>Reallocated</th>
<th>Ordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Practice</td>
<td>91%</td>
<td>9310</td>
<td>59897</td>
<td>6179</td>
<td>7046</td>
<td>74363</td>
</tr>
<tr>
<td>Proposed Model</td>
<td>95%</td>
<td>1957</td>
<td>62999</td>
<td>3265</td>
<td>7185</td>
<td>66437</td>
</tr>
</tbody>
</table>

2.2. Supply Chain Level Inventory Management

After the validation of the proposed RFID-based inventory management model, the presence of RFID at facility level is investigated in terms of its impact at the supply chain level. There are three business partners involved: material supplier, materials distributor, and manufacturing site. The materials supplier produces time-sensitive items based on mass production strategy; the larger the order quantities, the better for the supplier. The materials distributor tracks the orders at the manufacturing site and replenishes them twice a week. The value stream map for the current practice at the supply chain level includes 54 business processes. These processes involve a variety of non-value added activities, such as coding forms by hand, data entry using computer keyboard, regrouping inventory several times, sorting, error checking, data verification and reconciliation.

The current value stream map is analyzed to determine non-value added activities. Then, a second sub-group out of the non-value added group is selected by considering “what if RFID was deployed”. A thorough investigation of each step, along with interviews with experts and additional data collection, the RFID-enhanced value stream map is reduced to 18 steps, as opposed to 54, which represents 66.7% improvement in number of steps. Such an improvement not only makes the supply chain more reliable but also reduces the overall lead time drastically, which is estimated to be a reduction of 60%.

The supply chain level study revealed that technologies, such as RFID, allow supply chain partners to redefine the rules of business and roles of participants. For instance, the RFID-enhanced supply chain operations lead to elimination of cross-docking and staging, which are carried out buy the materials distributor in the current (non-RFID) practice. Elimination of such non-value added activities reduces the involvement of the materials distributor, thus diminishes its business potential. If the further deployment of RFID technologies is considered, then the materials manufacturer can actually deliver items directly to the manufacturing site and the reason for existence of the materials distributor becomes questionable. Such an approach removes the middle man and leads to the “vendor managed inventory” concept where the materials manufacturer is authorized to manage inventories at the customer site and to make decisions, such as when and how much inventory to ship.

2.3. A Lean Perspective

As demonstrated by the case study, automatic identification (Auto-ID) technologies provide operational visibility that needs to be utilized effectively in order to create a meaningful return on investment. Auto-ID technologies generate voluminous data in split second continuously. Deployment of such technologies without integration of effective data collection and consolidation schemes, and without decision making schemes that are based on meaningful, processed data can only lead to a chaotic state that can be defined as a “data rich, information poor” environment.

RFID, as well as other technologies under the Auto-ID umbrella, should be considered as a process improvement enabler; a tool that has the potential to make a process leaner by reducing or eliminating non-value added data capture processes. However, by solely deploying the technology cannot create an orderly process out of a chaotic state. A lean systems perspective is essential to first carry out a value stream mapping activity to understand the processes, information and material flow, and existing decision-making schemes. After identifying the performance metrics, opportunities for improvement can be identified based on the current state value stream map. Some of these improvements can be realized by RFID and some might require non-technological solutions, such as re-designing the layout, re-directing the flow, or as simple as adding a white board to improve communication. Once RFID is considered as a viable solution, then it is essential to first determine the future state of the value stream map, and second develop a benchmarking platform to validate its effectiveness. The benchmarking platform could be a simulation environment, a laboratory for hardware-based testing, or on-site prototyping depending on the level of complexity.

RFID applications create terabytes of real-time data. Collecting such voluminous data at a very high speed, its processing and storage impose severe challenges on current data management and storage strategies. Timely
communication of RFID data through business processes among organizations in a supply chain provides potential for seamless integration. In this area, research is essential in order to identify the issues, develop alternative solutions, and conduct cost-benefit analysis. RFID systems are too complex to allow realistic models to be evaluated analytically. Therefore, simulation-based approaches are essential for modeling and analysis of such systems, where a simple closed-form analytical solution through the use of a mathematical model is not possible due to the highly stochastic natures of these systems. Due to the limitation and uncertainty of RFID technology, simulation models that simply focus on RFID data-based decision-making, in isolation from actual hardware, can provide only limited insights. Prototypes that integrated RFID hardware, as well as decision-making, are essential to generate more realistic research results.

3. HARDWARE-IN-THE-LOOP SIMULATION AT LABORATORY

In order to test the performance of potential RFID-based solutions both at facility and supply chain levels, the Manufacturing Systems and Automation (MSA) laboratory at the University of Texas at San Antonio (UTSA) and the Embedded Systems and Networking laboratory (ESNL) at the Missouri University of Science and Technology (MST) at Rolla have established a Web-enabled RFID testbed that connects two remote sites for testing and validation of concepts. The connection between two sites is facilitated by Omnitrol Networks’ edge servers. The Omnitrol edge servers form a network that delivers peer-to-peer business services and enables real-time collaboration among business partners. The testbed is currently equipped with three edge servers (Omnitrol); two at UTSA and one at MST. Both sites are also equipped with RFID-equipped dock doors, smart shelves with RFID tracking, and various handheld readers as shown in Figure 2.

![Figure 2: A distributed testbed for RFID applications](image)

The simulation capability of Omnitrol is used for developing a virtual company that generates virtual events and RFID data, and interacts with the other two Omnitrols and the equipment connected to them. This is the hardware-in-the-loop simulation concept that allows for introducing more complexity, thus reality, into supply chain scenarios that can be tested in a very short time. Such expedited testing is also be used to generate very high volume “simulated” data to overload the network and to evaluate communication protocols and networking schemes [10].
4. CHALLENGES

Auto-ID represents a broad category of technologies that are used to identify objects without human intervention. In general, Auto-ID technologies, advanced sensing capabilities, and recent developments in the area of mobile wireless ad hoc networking provide a potential to establish a data-rich supply network. Such technological capabilities provide real-time visibility of each single entity in the supply chain; nevertheless they can only be effective if the real-time data can be integrated into the necessary business processes, such as manufacturing execution, production planning, and scheduling systems for improved decision-making. Myopic approaches, such as using Auto-ID technologies for asset tracking but without integrating it with scheduling, lead to a disconnect among a variety of islands of information in a supply chain and do not improve the system performance. Therefore, there is a need for comprehensive data models that facilitate intelligent decision-making among the organizations in a supply chain.

RFID imposes challenges at three layers: 1. Physical Layer: It represents the application environment that is equipped with a variety of Auto-ID technologies, advanced sensing capabilities, hand-held tools, and mobile wireless networks to facilitate timely communication; 2. Decision-Making Layer: This layer consists of effective decision-making models founded on efficient data harvesting, processing, and sharing schemes so that the performance goals at the Physical Layer (i.e., the application environment) can be met; and 3. Networking Layer: handles collecting, scheduling, and routing of voluminous data, which provides timely data and information flow for the Decision-Making Layer. The literature shows that most studies focus either on manufacturing-specific decision-making (manufacturing engineering and industrial engineering) or on networking (electrical and computer engineering) in isolation from each other. In other words, the solutions provided in this area are focused only on a particular layer and are isolated from the other layers by making various assumptions, such as “the read-rates on RFID readers are 100%” or “no data packets are dropped at the networking layer”. Such studies fall short from being realistic or complete. Therefore, integrative architectures that tie performance metrics at various levels in a supply chain with networking-level routing and scheduling protocols are required.

In general, RFID adds a new level, which consists of fine-grained data, to the traditional research and development methodologies. This new level imposed by RFID necessitates more sophisticated schemes for process modeling and data management.

5. CONCLUSIONS

Potential benefits of RFID, when implemented properly, are as follows:

- **Instantaneous Operator-Free Data Entry and Monitoring:** RFID readers and antennas communicate voluminous RFID data in milliseconds and have the capability to scan multiple items simultaneously, which significantly facilitates automation of many supply chain processes that are typically labor-intensive.

- **Effective Use of Labor:** Since many repetitious tasks can be automated via RFID, labor can be used for more effective tasks.

- **Visibility:** RFID provides real-time visibility for products across the supply chain providing accurate and detailed information, which can be used to improve efficiency, productivity, and quality. In addition to product-level visibility, an organization can track its valuable assets by tagging them. RFID technology also provides benefits for product recalls.

- **Mobile Databases:** Active RFID tags can be used as mobile databases; such tags can be updated dynamically as parts move across the supply chain.

There are four essential tasks that need to be executed in an integrated manner for a successful RFID implementation [5]:

- **A comprehensive analysis,** similar to a SWOT (Strengths-Weaknesses-Opportunities-Threats) analysis, of the work and information flow, existing business rules and practices, and current decision-making models used by the organization in the supply chain must be carried out. Such an analysis will reveal the dark spots in the processes and help prioritize the areas that can benefit from RFID.

- **Based on the results of the comprehensive analysis,** a technical feasibility study that investigates the integration of not only RFID but other technologies in those high-priority areas must be conducted. The technical feasibility study must focus on the following: (a) What information must be made available to
Radio Frequency Identification (RFID) in Manufacturing Enterprises: From Data to Information

make the processes more efficient and visible? (b) What are the potential interference and incompatibility issues associated with the RFID and other sensor technologies, including RFID middleware and other supplementary software?, and (c) What kind of a networking infrastructure, along with data routing and scheduling schemes, is essential to facilitate the necessary information flow in the most reliable and efficient way?

- An integrated product-process-system (re)design approach must be adopted in order to investigate, design, develop, and implement new business processes that rely on RFID and other sensor technologies. At the product level, proper tag placement alternatives that improve read rates must be investigated. At the process level, location of RFID antennas/readers must be properly determined such that the interference among antennas is minimized. At the system level, flow of tagged items must be synchronized with RFID tag reading frequencies so that the right information at the right time at the right location is captured and utilized to improve the overall performance of the organization. Such an integrated product-process-system (re)design approach must present an in-depth analysis of performance measures, such as cost savings, in order to yield viable alternative solutions for the organization.

- Pilot implementations in actual production and logistics environments must be carried out in order to verify the comprehensive and technical feasibility analyses and to validate the integrated product-process-system (re)design. These four steps require very strong managerial commitment in order to make the whole effort worthwhile for a company. “Islands of RFID applications” developed within departments but lacking integration at the company level will lead to waste of resources. In order to achieve an integrated RFID implementation first company-wide then across the supply chain, the main driver must be a strategic plan. Therefore, RFID implementation is more of a managerial decision than a technical one.

RFID is an enabling technology; it does not automatically bring business solutions to a supply chain. RFID technology simply facilitates visibility in a process. Tools, standards, and roadmaps that lead to effective utilization of such visibility to improve performance, reduce cost, and expedite decision-making are crucial for a successful RFID implementation.

6. SUMMARY

This paper presents a two-stage RFID deployment with emphasis on how non-value added activities can be eliminated by new technologies. First, at the facility level, a model for RFID-based inventory tracking is presented and benchmarked against the current practice. Second, a supply chain level perspective, which includes three business partners, is described. Advantages of RFID-based tracking at supply chain level are presented from a lean systems standpoint. Technical challenges of RFID implementations and how they can be overcome by laboratory-level hardware-in-the-loop simulations are also presented.

REFERENCES


Dynamic Analysis of a Simplified Centralised Supply Chain

Ian French\textsuperscript{1*}, Petia Sice\textsuperscript{2}, Mu Niu\textsuperscript{2}

\textsuperscript{1}School of Science and Technology  \hspace{1cm} \textsuperscript{2}School of Computing, Engineering and IS
\textsupersette{University of Teesside}  \hspace{1cm} \textsupersette{Northumbria University}
\textsupersette{Middlesbrough, UK}  \hspace{1cm} \textsupersette{Newcastle, UK}

\textbf{ABSTRACT}

This paper presents analysis of the behaviour of a model of a simplified centralised supply chain. The research was conducted within the manufacturing sector and involved a local manufacture of safety equipment. A simplified model of the companies centralised supply chain has been developed and validated. Simulation and analysis has been performed using non-linear dynamics and control theory. The findings suggest that the current management structure may result in the generation of destructive oscillations of inventory and that managers at the production unit (if acting in isolation) will be unable to correct the situation.

1. INTRODUCTION

‘The Company’ is an international corporation, manufacturing safety equipment. It has a presence in approximately 200 countries, on all continents. ‘The Company’ is equipped with a sophisticated MRP environment for sharing information and enhancing communication between production and marketing. However, the performance of the supply chain is still perceived as problematic and it is not unusual for high inventories to mount up, or backlogs to develop; causing sever disruption as the production unit is forced to deal with the crisis to overcome them. As a result, a centralised supply chain system was recently implemented with the purpose of diminishing costs and avoiding backlogs. This, however, made the production unit’s planning managers even more worried as it was difficult to predict what the consequences of centralised inventories would be for the manufacturing plant. A team consisting of the planning department and researchers from Northumbria University was formed to study the impact of different decision strategies on the behaviour of the supply chain. The purpose of the study was to model the material and information flows (including the decision processes), and, to provide explanation and insight into the link between decision strategies and inventory and backlog oscillations.

2. THE STRUCTURE OF THE SUPPLY CHAIN

Like many other manufactures ‘The Company’ makes use of a cascaded distribution system, with inventory holders at several different levels. In its simplest form (Figure 1), the central Hub, based in Germany, receives stock from the UK Factory and ships this to the main EU markets and to the Pacific and US Hubs. Each of the Hubs acts as a business unit and maintains a warehouse.

![Figure 1: Schematic representation of the centralised supply chain proposed by ‘The Company’](image-url)
The Hubs receive orders from the sales agencies linked to them. The sales agencies do not keep inventory or receive shipments from the Hub. The role of the agencies is to receive orders from the customer and to make sales forecasts for their area. They then forward the orders and the forecasts to their respective Hub. The Hubs ship the goods directly to the customer and communicate with production units to help establish production requirements.

The relationships that exist between the Central Hub and the Secondary Hubs are similar to those that exist between the Factory and the Central Hub. Moreover, the research was conducted at the production unit (Factory). Thus, to better understand the impact of the supply chain structure on the factory, a simplified system dynamics model was constructed representing the relationships between the Factory and the Central Hub. The boundaries of the model were chosen based on discussions with company staff. These determined that the model should include the customer, the production plant, the Hub in Germany and the relevant decision making processes. The structure of this portion of the supply chain is illustrated in Figure 2.

![Schematic diagram showing relationships between factory and central hub.](image)

**Figure 2:** Schematic diagram showing relationships between factory and central hub.

### 2. The Model

In the model, which is discrete and based loosely on that of the Beer Game [1, 2], inventories are updated (using a 1 month sample interval) by adding incoming shipments and subtracting outgoing shipments.

\[
H_{inv}(t) = \max(0, H_{inv}(t-1) + F_{ship}(t-1) - H_{ship}(t)) \tag{1}
\]

\[
F_{inv}(t) = \max(0, F_{inv}(t-1) + F_{prod}(t-1) - F_{ship}(t)) \tag{2}
\]

To the extent that inventory plus incoming shipments are sufficient, outgoing shipments are existing backlog plus incoming orders. Otherwise, outgoing shipments are incoming shipments plus inventory, and the new inventory is empty.

\[
H_{ship}(t) = \min(H_{orders}(t) + H_{blk}(t-1), H_{inv}(t-1) + F_{ship}(t-1)) \tag{3}
\]

\[
F_{ship}(t) = \min(H_{req}(t+1) + F_{blk}(t-1), F_{inv}(t-1) + F_{prod}(t-1)) \tag{4}
\]

In a similar way, backlogs are updated by adding incoming orders and subtracting outgoing shipments. If incoming orders plus backlogs are completely covered by incoming shipments plus existing inventory the new backlog is empty.

\[
H_{blk}(t) = \max(0, H_{blk}(t-1) + H_{orders}(t) - (H_{inv}(t-1) + F_{ship}(t-1))) \tag{5}
\]

\[
F_{blk}(t) = \max(0, F_{blk}(t-1) + H_{req}(t+1) - (F_{inv}(t-1) + F_{prod}(t-1))) \tag{6}
\]

The management policies that are used to generate Hub orders and Factory production are based on an anchoring and adjustment heuristic [1]. The HUB forecast forms the main driver for the ordering...
policy. Decision makers operate a form of adaptive forecast, which is formed from incoming orders based on a simple first-order exponential prediction. The mechanism underlying this forecast is characterised by an expression, which takes the form:

\[ H_{\text{forecast}}(t+2) = (1 - \theta) H_{\text{orders}}(t) + \theta H_{\text{forecast}}(t+1) \]  

(7)

Here, \( \theta \) represents the rate at which the forecast is adapted. \( \theta = 1 \), corresponds to a constant forecast, and \( \theta = 0 \) describes a situation where the immediately previous value of HUB orders is used as an estimate of future demand. The process is based on a 2 month forecast. That is, based on the current trends in orders and stock levels, the Hub forecasts its requirements for 2 months ahead. The factory then produces this in the current month so that it can be shipped to the Hub next month; in time to meet the expected demand.

In addition, decision makers were found to adjust orders above or below the expected demand in order to keep their inventory and supply line at the desired level. Thus, the overall ordering policy (HUB request & Factory production) may be characterised by the expressions:

\[ H_{\text{req}}(t+2) = \max(0, \alpha (Q - H_{\text{inv}}(t) + H_{\text{blk}}(t)) \]  

\[ - \alpha \beta (F_{\text{blk}}(t) + F_{\text{ship}}(t)) + H_{\text{forecast}}(t+2)); \]  

(8)

\[ F_{\text{prod}}(t) = \max(0, \alpha (Q - H_{\text{inv}}(t) + H_{\text{blk}}(t)) + H_{\text{req}}(t+2)); \]  

(9)

where, \( \alpha \), (which lies in the range 0-1) defined the fraction of the discrepancy between the desired inventory \( Q \) and the effective inventory (Inventory – Backlog), that need to be ordered each month. \( \beta \) (which again lies in the range between \( 0 - 1 \)) is the fraction of the supply line (requests or production already ordered but not yet received) that is taken into account. If \( \beta = 1 \), decision makers fully recognise shipments already in the pipeline and do not double order. If \( \beta = 0 \), requests (orders) are immediately forgotten and new requests to cover the same lack of inventory are placed in the next month.

3. Model Analysis

An inspection of the model equations (Equations 1 to 9) indicates that the primary nonlinearities present in the model are of a saturation form. Because of this, initial model analysis is based on a linear small-signal representation, in which all backlogs are assumed to be zero. Such an analysis will allow the soundness (or otherwise) of the basic management (feedback) structure to be assessed and will also provide an insight into the various transitions towards instability, without this information being lost in the added complications introduced by nonlinear behaviour.

Thus, if we define the general discrete state space representation as:

\[ X(k) = A X(k-1) + B U(k) \]  

(10)

\[ Y(k) = C X(k) \]

the state vector \( X(k) \) as:

\[
X(k) = \begin{bmatrix}
H_{\text{ship}}(t) \\
H_{\text{inv}}(t) \\
H_{\text{req}}(t + 2) \\
H_{\text{forecast}}(t + 2) \\
F_{\text{ship}}(t) \\
F_{\text{inv}}(t) \\
F_{\text{prod}}(t)
\end{bmatrix}
\]  

(11)
and, the input $U(k)$ as:

$$U(k) = \begin{bmatrix} H_{\text{orders}}(t) \\ Q_H(t) \\ Q_F(t) \end{bmatrix}$$ (12)

Then, it is easily shown that the system matrices can be written as:

$$A =\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & -\alpha_H & -\alpha_H\beta_H & 0 & -\alpha_H & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 1 \\ 0 & -\alpha_H & \alpha_F & -\alpha_H\beta_H & 0 & -\alpha_H & -\alpha_F & -\alpha_F \end{bmatrix}$$ (13)

$$B = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1+\alpha_H - \theta & \alpha_H & 0 \\ 1-\theta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1+\alpha_H - \theta & \alpha_H \alpha_F \end{bmatrix} : C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$ (13)

A Simulink block diagram of this linearised model is presented in Figure 3.

In linear system analysis the relationship between behavioural modes and the location of the eigenvalues of $A$ is well established. Consequently, it is clear that a plot of the trajectories of the eigenvalues in response to the variation of a specific parameter is a good indication of the influence that the parameter has on system behaviour. Inspection of Equation 7 indicates that the $A$ matrix contains three variable parameters: $\alpha$, $\beta$ and $\theta$. In studies of similar supply chains [2] it is generally accepted that for the majority of managers the value of $\theta$ tends to remain constant. This is confirmed by finding at ‘The Company’, where conversations with managers suggest that forecast is amended with a smoothing time of approximately 4 months ($\theta = 0.75$). In addition, inspection of the block diagram (Figure 3) indicates that $\theta$ only exists in the first order input prefilter and, thus, is outside of the feedback structure. $\alpha$ and $\beta$, on the other hand clearly lie inside the feedback loops and consequently will significantly influence the dynamic behaviour generated by the system itself. For these reasons, in the study presented here, $\theta$ will remain fixed at 0.75 and the analysis will concentrate only on variations in $\alpha$ and $\beta$.

### 3.1. The Influence of $\alpha$ & $\beta$

Figure 4 shows a plot of the eigenvalue locus of the $A$ matrix, calculated as the solution to $\det(zI - A) = 0$ (where $I$ is an identity matrix of dimension $A$): a) when $\beta = 0$ and $\alpha$ is varied from 0 to 1; and: b) when $\alpha = 1$ and $\beta$ is varied from 0 to 1. A quick inspection of the figure suggests that there is (at
least) one pair of complex eigenvalues and that these intersect with the unit circle at a position 0.5 ± j0.866, for the condition $\alpha = 1$, $\beta = 0$. Clearly, $\alpha$ has a destabilising influence, since the complex eigenvalue pair (Figure 4a) move toward the edge of the unit disc as $\alpha$ increases. $\beta$, on the other hand, is clearly stabilising, since the eigenvalues (Figure 4b) move toward the centre of the unit disc as $\beta$ increases.

Figure 3: Small signal block diagram representation of the linearised model of the simplified supply chain.

Figure 4: Locus of eigenvalues of $A$ plotted on the $z$-plane for $\theta = 0.75$. (a) $\beta = 0$ and $\alpha$ varied from 0 to 1. (b) $\alpha = 1$ and $\beta$ varied from 0 to 1. Clearly, the complex pair migrates toward the edge of the unit circle as $\alpha$ increases, while they move toward the centre of the unit circle with increasing $\beta$. 
To better understand the observed behaviour, we need to re-examine the structure of the system block diagram (Figure 3). Inspection of the figure reveals that the Factory and the Hub have ‘isolated’ feedback structures. Therefore, by considering their individual characteristic equations, these can be analysed independently.

Thus, for the condition $\beta = 0$ (depicted in Figure 4a), the Factory characteristic equation is:

$$ (z + \alpha)(z - 1) + \alpha = 0 \quad (9) $$

This has two eigenvalues, one at $z = 0$ and a second, which is always real and which lies in the range:

$$ z = 1 \rightarrow 0 \text{ as } \alpha = 0 \rightarrow 1. $$

The Hub characteristic equation is:

$$ (z^2 + \alpha)(z - 1) + \alpha = 0 \quad (10) $$

This has three eigenvalues. Again one of these is at $z = 0$, the other two form a second order pair that become complex when $\alpha > 0.25$. It is this pair that are clearly identified in Figure 4a. Moreover, it is the Hub’s dynamics and not the Factory’s that are the potential source of unstable behaviour. The Hub, potentially, becoming unstable for any value of $\alpha > 1$, whilst the Factory would be stable for any value of $\alpha < 2$.

4. DISCUSSION

In the model, management decisions are represented by the 3 corrective (negative) feedback loops. An initial estimate of Factory production is made based on the forecast for Hub sales in 2 months time. This estimate is then corrected (feedback 1) based on the discrepancy between the current Hub inventory and its desired level multiplied by a ‘correction factor’, $\alpha$. This modified estimate is then re-corrected (feedback 2) to take into account any previously placed Hub orders (based on previous attempts to correct the Hub inventory discrepancy) that have not yet arrived and are thus ‘still in the pipeline; in this case the multiplying factor is $\beta$. The final correction to Factory production (feedback 3), is made based on the discrepancy between current Factory inventory and its desired level, once again this is multiplied by $\alpha$. In this form, the ‘management policy’ appears to be, primarily; one of stock control in the presence of external disturbances (actual product sales), and, the feedback structure that is in place bares a strong similarity to the classical ‘cascade control’ structure, familiar to many engineers.

Within the context of this specified ‘management policy’ the two correction factors $\alpha$ and $\beta$, can be thought of as representing the psyche of the decision maker: $\alpha = 1$ and $\beta = 0$, represents one extreme; an aggressive manager thinking only within the confines of his area of control and not recognising the impact of previous decisions (no account of history). In simulations studies this resulted in persistent oscillations, generated purely by the decision making strategy and not as the result of some external influence (as is often claimed by managers with such tendencies). $\beta > 0$, represents a manager that thinks outside of immediate area of control and takes some or full account of history. Taking into account the impact of previous decisions, has a stabilising effect. $\alpha = 1$, $\beta = 1$ would seem to represent the ideal case (just in time, or deadbeat, management – with the eigenvalues lying at the origin of the z-plane; Figure 4). It is well known, however, that such control policies often prove unreliable, as they tend to be intolerant of unexpected events (such as unforeseen production hold-ups – this effect will be demonstrated during the oral presentation).

Without doubt, however, the most revealing insight presented here, is that the current supply chain management policies have effectively decoupled of the Factory and Hub feedback structures. In the worst case this may lead to the generation of destructive oscillations of inventory. The cause would be a Hub manager who corrects for the current inventory deficit and then adds a little for good luck. This would result in $\alpha > 1$ and the Hub eigenvalues would lie outside the unit circle. Clearly, under such circumstances the decoupling of the Factory and Hub feedback structures means that even the most competent Factor manager (if acting in isolation), would be unable to mitigate the effect. Moreover, it will be shown in the oral presentation that this effect worsens significantly if hold-ups occur in production. Thus, good management personnel (and policies) are critical if significant problems are to be avoided with the proposed structure.
5. CONCLUSION

The analysis presented here represents only the initial stages of an investigation into behaviour of the proposed supply chain. Never-the-less this analysis has found favour within ‘The Company’ and has proven itself to be a powerful decision support tool.

It has become obvious that in our investigations so far we have only just scratched the surface of the rich variety of possible behaviours that can emerge. It is the intention, therefore, to continue this research and to further investigate:

- The influence of errors in the forecast.
- The effect of using independent decision strategies (having different values of $\alpha$ and $\beta$ at the Factory and the Hub) at different levels in the supply chain.
- The impact of the secondary Hubs.
- Modifying the decision strategies (perhaps to a full state variable feedback form) to take into account better information flow.

REFERENCES


# AUTHOR INDEX

## A
- Ab Raham, M. N - 469
- Abbas, Ahmed - 45
- Abdullah, Sajid - 824
- Abdullah, S. - 1155
- Abedinzadeh, Reza - 850
- Abolhasani, Hooman - 873
- Ackermann, Jörg - 1512
- Adeoye, Adeyinka - 1277
- Adhikari, Diwas - 777
- Afshouk, Abdulmunam - 816
- Aggogeri, Francesco - 1178
- Ahmad, S. - 838
- Ahmad, Munir - 67
- Ali, Zulf - 14
- Al-Bazi, Ammar - 1340
- Al-Kindi, Nabil Said - 656
- Alvareng, Esther - 1450
- Alvareng, Esther - 1467
- Amelia, Lily - 1124
- Amelia, Lily - 1155
- Amin, S.Y.M - 838
- Andersson, Paul - 964
- Andersson, Paul - 1269
- Ansell, M. P. - 947
- Arana Arexolaleiba, N - 477
- Arezoo, Behrooz - 112
- Ashayeri, Jalal - 1073
- Ashayeri, Jalal - 1481
- Assadi, G. - 1101
- Auger, Vincent - 14
- Aurich, J.C. - 1309
- Azahari, Che Husna - 971
- Azzopardi, Sandro - 407

## B
- Bachlaus, M. - 1458
- Baciou, George - 89
- Badak, Zeynel - 1386
- Baek, Jae Yong - 202
- Baek, Jun-Geol - 415
- Baek, Jun-Geol - 598
- Baradaran, Siamak - 664
- Barber, K. - 1251
- Barrans, Simon - 919
- Bashir, Anwar - 1084
- Bayraktar, Erkan - 319
- Behnamian, J. - 493
- Behnamian, J. - 501
- Bellgran, Monica - 751
- Bertsche, Bernd - 1286
- Bi, Z.M. - 360
- Birger Stav, John - 939
- Borg, Jonathan C - 399
- Boyce, Michael - 37
- Boyce, Michael - 45
- Bradely, Edward - 1058
- Bridgwood, Kirk - 887
- Bridgwood, Kirk - 995
- Brisbon, Brandon - 1163
- Burghardt, Thomas - 1427
- Cannizzarro, L. - 1042
- Cantoni, S. - 1356
- Cao, Wenping - 218
- Cao, Wenping - 656
- Caricato, Pierpaolo - 605
- Carrino, L. - 1356
- Certa, A - 1026
- Certa, A - 1034
- Chang, Xiaomeng - 1163
- Che Haron, C.H. - 1124
- Chen, Zhangjin - 680
- Chen, Zhangjin - 686
- Chen, F. Frank - 429
- Chen, F. Frank - 1458
- Chen, F. Frank - 1476
- Chen, Hongyi - 613
- Chiado, Joseph - 210
- Choi, Hung-Won - 194
- Choi, Hun Zong - 202
- Cissé, Adji - 800
- Connolly, Shahrazad - 291
- Connolly, Stephen N - 291
- Connor, Christopher - 23
- Coole, Tim - 734
- Copeland, Keith - 2
- Crabtree, P. - 947
- Crostack, Horst-Artur - 1192
- Crostack, Horst-Artur - 1218
- Curcurù, Giuseppe - 1019
- de la Calle, Alberto - 1450
- De Vin, Leo - 1317
- De Vin, Leo - 1348
- Dhafr, Nasreddin - 37
- Dhafr, Nasreddin - 45
- Dholes, Vimal G. - 947
- Dias, Altamir - 130
- Dias, Altamir - 436
- Dias, Altamir - 1139
- Diaz, Fernando - 1467
- Dietmair, Anton - 226
- Dongil, Ming - 1476
- Doolen, Toni - 1226
- dos Santos, Marcelo T. - 122
- Dossou, Paul Eric - 1442
- Doverborn, Josefine - 1348
- Dukovski, Vladimir - 97
- Dvorakova, Lilia - 179
- Dvoráková, Lilia - 257
- Eberspaecher, Philipp - 226
- Eguta, Ignacio - 621
- Ekren, Banu - 376
- El Barky, Sahar - 540
- El Sayed, Azziz - 540
- Elahi, Maudood - 282
- Enea, M. - 1042
- English, John - 1186
- Ericsson, Dag - 1169
- Falivene, M. - 1411
- Fang, Mike J. - 1091
- Farris, Jennifer - 1226
- Farsi, Mohamad Ali - 112
- Fatemi Ghomi, S.M.T - 485
- Fatemi Ghomi, S.M.T - 493
- Fatemi Ghomi, S.M.T - 501
- Fatemi Ghomi, S.M.T - 509
- Fatemi Ghomi, S.M.T - 516
- Fatemi Ghomi, S.M.T - 549
- Fatemi Ghomi, S.M.T - 557
- Fatemi Ghomi, S.M.T - 664
- Fatemi Ghomi, S.M.T - 672
- Feng, His-Yung - 186
- Ferreira, Joao - 122
- Ferreira, Joao - 311
- Ferreira, Joao - 1200
- Fertsch, Marek - 532
<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fertsch, Marek</td>
<td>1302</td>
</tr>
<tr>
<td>Fillatreau, P</td>
<td>477</td>
</tr>
<tr>
<td>Foit, Krzysztof</td>
<td>423</td>
</tr>
<tr>
<td>Ford, David</td>
<td>452</td>
</tr>
<tr>
<td>Forrester, Paul</td>
<td>786</td>
</tr>
<tr>
<td>Forrester, Paul</td>
<td>1234</td>
</tr>
<tr>
<td>French, Ian</td>
<td>1365</td>
</tr>
<tr>
<td>French, Ian</td>
<td>1527</td>
</tr>
<tr>
<td>Fruggiero, Fabio</td>
<td>147</td>
</tr>
<tr>
<td>Fruggiero, Fabio</td>
<td>1333</td>
</tr>
<tr>
<td>Galante, G</td>
<td>1019</td>
</tr>
<tr>
<td>Galante, G</td>
<td>1026</td>
</tr>
<tr>
<td>Galante, G</td>
<td>1034</td>
</tr>
<tr>
<td>Gallagher, Mike</td>
<td>242</td>
</tr>
<tr>
<td>Garrido Campos, J.</td>
<td>792</td>
</tr>
<tr>
<td>Garrido Campos, J.</td>
<td>896</td>
</tr>
<tr>
<td>Garza-Reyes, J.A.</td>
<td>786</td>
</tr>
<tr>
<td>Garza-Reyes, J.A.</td>
<td>1234</td>
</tr>
<tr>
<td>Gatsou, Anna</td>
<td>1309</td>
</tr>
<tr>
<td>Gaughran, William</td>
<td>234</td>
</tr>
<tr>
<td>Gaughran, William</td>
<td>242</td>
</tr>
<tr>
<td>Gaughran, William</td>
<td>808</td>
</tr>
<tr>
<td>Geraghty, John</td>
<td>743</td>
</tr>
<tr>
<td>Geraghty, John</td>
<td>1294</td>
</tr>
<tr>
<td>Gerrard, A.M.</td>
<td>250</td>
</tr>
<tr>
<td>Ghandehary, Mahsa</td>
<td>850</td>
</tr>
<tr>
<td>Ghani, Jaharath A.</td>
<td>469</td>
</tr>
<tr>
<td>Ghasemi Yaghin, R.</td>
<td>485</td>
</tr>
<tr>
<td>Ghasemi Yaghin, R.</td>
<td>549</td>
</tr>
<tr>
<td>Ghasemi Yaghin, R.</td>
<td>672</td>
</tr>
<tr>
<td>Gibbons, Paul</td>
<td>295</td>
</tr>
<tr>
<td>Gilani, Sadri</td>
<td>75</td>
</tr>
<tr>
<td>Gilani, Parisa</td>
<td>60</td>
</tr>
<tr>
<td>Gillick, Christy</td>
<td>912</td>
</tr>
<tr>
<td>Gogouvitis, X.V.</td>
<td>1309</td>
</tr>
<tr>
<td>Gohareh, M. Morady</td>
<td>549</td>
</tr>
<tr>
<td>Golinska, Paulina</td>
<td>532</td>
</tr>
<tr>
<td>Golnabi, Hossein</td>
<td>343</td>
</tr>
<tr>
<td>Golnabi, Hossein</td>
<td>368</td>
</tr>
<tr>
<td>Goss, Raymond</td>
<td>777</td>
</tr>
<tr>
<td>Grabowski, Cezary</td>
<td>155</td>
</tr>
<tr>
<td>Greco, Antonio</td>
<td>605</td>
</tr>
<tr>
<td>Guerrero, Fernando</td>
<td>621</td>
</tr>
<tr>
<td>Hachek, Sascha</td>
<td>1427</td>
</tr>
<tr>
<td>Haghighatzadeh, A.</td>
<td>343</td>
</tr>
<tr>
<td>Haghighatzadeh, A.</td>
<td>368</td>
</tr>
<tr>
<td>Haider, Julfiyar</td>
<td>858</td>
</tr>
<tr>
<td>Hashemin, S.S.</td>
<td>664</td>
</tr>
<tr>
<td>Hashmi, M. S. J.</td>
<td>1294</td>
</tr>
<tr>
<td>Hedenstierna, P.</td>
<td>573</td>
</tr>
<tr>
<td>Hedenstierna, Philip</td>
<td>1169</td>
</tr>
<tr>
<td>Heo, Eun-Young</td>
<td>429</td>
</tr>
<tr>
<td>Heo, Eun-Young</td>
<td>904</td>
</tr>
<tr>
<td>Heragu, Sunderesh</td>
<td>376</td>
</tr>
<tr>
<td>Higgins, Stan</td>
<td>4</td>
</tr>
<tr>
<td>Hilletofth, Per</td>
<td>573</td>
</tr>
<tr>
<td>Hilletofth, Per</td>
<td>1169</td>
</tr>
<tr>
<td>Hillier, Graham</td>
<td>1</td>
</tr>
<tr>
<td>Hilmola, Olli-Pekka</td>
<td>573</td>
</tr>
<tr>
<td>Hilmola, Olli-Pekka</td>
<td>1169</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>23</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>29</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>351</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>452</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>833</td>
</tr>
<tr>
<td>Hodgson, Simon</td>
<td>1375</td>
</tr>
<tr>
<td>Holm, Magnus</td>
<td>1348</td>
</tr>
<tr>
<td>Horbach, Sebastian</td>
<td>1512</td>
</tr>
<tr>
<td>Hossain, Syed Shahadat</td>
<td>282</td>
</tr>
<tr>
<td>Hossien, Nazir</td>
<td>1208</td>
</tr>
<tr>
<td>Hsieh, Fu-Shiung</td>
<td>717</td>
</tr>
<tr>
<td>Hsieh, Fu-Shiung</td>
<td>769</td>
</tr>
<tr>
<td>Huang, George</td>
<td>3</td>
</tr>
<tr>
<td>Huang, George</td>
<td>1091</td>
</tr>
<tr>
<td>Hughes, David</td>
<td>29</td>
</tr>
<tr>
<td>Huscika, Kyle</td>
<td>1186</td>
</tr>
<tr>
<td>Hwang, Jibin</td>
<td>598</td>
</tr>
<tr>
<td>Iannone, R.</td>
<td>1411</td>
</tr>
<tr>
<td>Ibrahim, M.H.I.</td>
<td>838</td>
</tr>
<tr>
<td>Ijomah, Winifred</td>
<td>210</td>
</tr>
<tr>
<td>Imani Asrai, R.</td>
<td>955</td>
</tr>
<tr>
<td>Ion, William</td>
<td>540</td>
</tr>
<tr>
<td>Iranpoor, Mehdi</td>
<td>672</td>
</tr>
<tr>
<td>Ismail, A.R.</td>
<td>274</td>
</tr>
<tr>
<td>Ismail, A.R.</td>
<td>469</td>
</tr>
<tr>
<td>Ismail, A.R.</td>
<td>971</td>
</tr>
<tr>
<td>Ismail, A.R.</td>
<td>1124</td>
</tr>
<tr>
<td>Izagirre, A.</td>
<td>477</td>
</tr>
<tr>
<td>Jafari, R.</td>
<td>343</td>
</tr>
<tr>
<td>Jahn, Hendrick</td>
<td>1427</td>
</tr>
<tr>
<td>Jaksic, Nebojsa</td>
<td>444</td>
</tr>
<tr>
<td>Jali, E.</td>
<td>343</td>
</tr>
<tr>
<td>Jamaludin, K. R.</td>
<td>838</td>
</tr>
<tr>
<td>Janardhan, G.Ranga</td>
<td>1419</td>
</tr>
<tr>
<td>Jasim, Abbas</td>
<td>833</td>
</tr>
<tr>
<td>Jolai, F.</td>
<td>516</td>
</tr>
<tr>
<td>Jung, So Young</td>
<td>202</td>
</tr>
<tr>
<td>Kadlecova, Tereza</td>
<td>179</td>
</tr>
<tr>
<td>Kalinowski, Krzysztof</td>
<td>629</td>
</tr>
<tr>
<td>Karimian, Mehdi</td>
<td>850</td>
</tr>
<tr>
<td>Khalid, K.</td>
<td>1155</td>
</tr>
<tr>
<td>Khamis, Nor Kamaliana</td>
<td>274</td>
</tr>
<tr>
<td>Khan, Fahd N.</td>
<td>865</td>
</tr>
<tr>
<td>Kim, Bo Hyun</td>
<td>202</td>
</tr>
<tr>
<td>Kim, Dong-Won</td>
<td>429</td>
</tr>
<tr>
<td>Kim, Dong-Won</td>
<td>904</td>
</tr>
<tr>
<td>Kim, Dong-Won</td>
<td>1458</td>
</tr>
<tr>
<td>Kim, Dongil</td>
<td>598</td>
</tr>
<tr>
<td>Kim, Seong-Jin</td>
<td>429</td>
</tr>
<tr>
<td>Kim, Sung-Shick</td>
<td>415</td>
</tr>
<tr>
<td>Kim, Sung-Shick</td>
<td>598</td>
</tr>
<tr>
<td>Klemmt, Andreas</td>
<td>1325</td>
</tr>
<tr>
<td>Kononis, Athanasios</td>
<td>81</td>
</tr>
<tr>
<td>Kononis, Athanasios</td>
<td>1050</td>
</tr>
<tr>
<td>Klute, Sandra</td>
<td>1192</td>
</tr>
<tr>
<td>Kontolatis, Nikolaos</td>
<td>139</td>
</tr>
<tr>
<td>Kost, G</td>
<td>336</td>
</tr>
<tr>
<td>Krahtov, Konstantin</td>
<td>461</td>
</tr>
<tr>
<td>Kretz, Daniel</td>
<td>105</td>
</tr>
<tr>
<td>Krishnamurthy, Ananth</td>
<td>376</td>
</tr>
<tr>
<td>Kubasad, Sachin</td>
<td>581</td>
</tr>
<tr>
<td>Kumile, Chris Matlalo</td>
<td>881</td>
</tr>
<tr>
<td>Kyriakopoulos, Kostas</td>
<td>139</td>
</tr>
<tr>
<td>L</td>
<td></td>
</tr>
<tr>
<td>La Fata, C. Manuela</td>
<td>1042</td>
</tr>
<tr>
<td>Lambiasi, Alfredo</td>
<td>147</td>
</tr>
<tr>
<td>Lambiasi, Alfredo</td>
<td>1333</td>
</tr>
<tr>
<td>Lange, Janis</td>
<td>1325</td>
</tr>
<tr>
<td>Larrinaga, Miguel A.</td>
<td>1467</td>
</tr>
<tr>
<td>Leahy, Keelin</td>
<td>808</td>
</tr>
<tr>
<td>Lee, Chan-Gie</td>
<td>429</td>
</tr>
<tr>
<td>Lee, Hee-Kwan</td>
<td>904</td>
</tr>
<tr>
<td>Lee, Ivan</td>
<td>700</td>
</tr>
<tr>
<td>Lee, Ivan</td>
<td>709</td>
</tr>
<tr>
<td>Lee, Jong-Yeong</td>
<td>904</td>
</tr>
<tr>
<td>Lee, Robert</td>
<td>37</td>
</tr>
<tr>
<td>Lee, Robert</td>
<td>45</td>
</tr>
<tr>
<td>Lee, Seok Woo</td>
<td>202</td>
</tr>
<tr>
<td>Lemke, Sascha</td>
<td>1427</td>
</tr>
<tr>
<td>Lewis, Huw</td>
<td>979</td>
</tr>
<tr>
<td>Li, Xuekun</td>
<td>6</td>
</tr>
<tr>
<td>Li, W.J.</td>
<td>1491</td>
</tr>
<tr>
<td>Li, Yanfeng</td>
<td>1163</td>
</tr>
<tr>
<td>Lindke, Richard</td>
<td>613</td>
</tr>
<tr>
<td>Little, James</td>
<td>743</td>
</tr>
<tr>
<td>Liu, Wen-Hsing</td>
<td>1226</td>
</tr>
<tr>
<td>Liu, R.</td>
<td>163</td>
</tr>
<tr>
<td>Liu, W.N.</td>
<td>1091</td>
</tr>
<tr>
<td>Lozano, Sebastian</td>
<td>621</td>
</tr>
<tr>
<td>Lu, Joan</td>
<td>932</td>
</tr>
</tbody>
</table>
Lu, Joan - 939
Lu, Haili - 1091
Ludin, Asri - 274
Lupo, T. - 1026
Lupo, T. - 1034

M
Ma, Ji - 186
Maghool, Elham - 1109
Malmborg, Charles J - 376
Mander, Todd - 700
Mander, Todd - 709
Marjani, M.R. - 557
Martín, Ricardo Marín - 896
Martins, Daniel - 436
Matchett, Andrew - 700
Matchett, Andrew - 709
Matthew, Arun Tom - 1132
Mathis, Jonas - 1218
Maxim, Carmen - 777
Mazzola, Marco - 1178
Mckie, Martin - 887
Mckie, Martin - 995
McKie, Richard - 887
Md Deros, Baba - 274
Md Deros, Baba - 249
Md Deros, Baba - 1243
Militzer, Joerg - 105
Miller, John M. - 1186
Mirab, M - 516
Miranda, Salvatore - 1411
Mirian, Saeed - 425
Mitchell, Philip - 1442
Mohamed, Ayad K. - 1294
Mohandass, Gajendiran - 444
Moslehi, Ghasem - 590
Mpfu, Khumbulani - 881
Mráz, Luboš - 939
Muhamad, N - 469
Muhamad, N. - 838
Muhamad, N. - 873
Muhamad, N. - 971
Muhamad, N. - 1124
Müller, Egon - 1512

N
Nabhani, Farhad - 23
Nabhani, Farhad - 29
Nabhani, Farhad - 351
Nabhani, Farhad - 700
Nabhani, Farhad - 709
Nabhani, Farhad - 833
Nabhani, Farhad - 887
Nabhani, Farhad - 1050
Nabhani, Farhad - 1058
Nabhani, Farhad - 1084
Nabhani, Farhad - 1208
Nabhani, Farhad - 1365
Nabhani, Farhad - 1375
Naouli, Fatiha - 1435
Nasseri, A. - 163
Nasseri, A. - 955
Nasseri, A. - 1356
Nebel, Steffen - 1286
Neelakanteswara Rao, A - 1066
Nekolemehr, Nooshin - 590
Neumann, Tim - 105
Neumann, Tim - 1403
Newman, S.T. - 163
Newman, S.T. - 947
Newman, S.T. - 955
Newman, S.T. - 1091
Newman, S.T. - 1356
Ng, Amos - 1348
Niola, Vincenzo - 383
Niola, Vincenzo - 391
Niu, Mu - 1527
Nookabadi, Ali - 1395
Nopiah, Z. M. - 469
Nosoohi, Iman - 1395
Nucci, Francesco - 605
Nylund, Hasse - 964
Nylund, Hasse - 1269

O
Oduoza, Chike - 52
Oduoza, C.F. - 760
Oduoza, C.F. - 1251
O'Hare, William - 282
Ohman, Peter - 1116
Oleskow Szlapka, J. - 1302
Olley, P. - 1003
Oloomi, S. A. A. - 1257
O'Neill, James - 1011
Ospyiw, David - 734
Ovtcharova, Jivka - 461
Ovtcharova, Jivka - 987

P
Paca, J - 250
Pandilov, Zoran - 97
Park, Jonghyuck - 415
Park, Joohyuk - 971
Park, Hong-Seok - 194
Park, Jin-Woo - 194
Pasek, Zbigniew. J - 89
Pasirayi, Godfrey - 14
Passannanti, A. - 1034
Passannanti, G. - 1026
Passannanti, G. - 1034
Passannanti, G. - 1042
Patel, Chintan - 726
Patwary, Masum A - 282
Pawlewski, Pawel - 532
Peacock, Tim - 786
Peacock, Tim - 1234
Pein, Raoul Pascal - 939
Pellegrini, Nicola - 1178
Permsal, Joakim - 1116
Persson, Martin - 865
Phealan, Pat - 242
Pinedo-Cuenca, R - 81
Pop, R - 477

Q
Qi, H.S. - 1003
Quinn, Sonya - 234

R
Racero, Jesus - 621
Rahman, M. N. Ab. - 838
Rahman, M. N. Ab. - 1243
Rahman, Norashikin - 1243
Rahman, Pattanathu - 14
Rahman, S. A. - 469
Rama Krishna, L.Siva - 1419
Ramli, R. - 524
Ramli, R. - 1124
Ran, Feng - 680
Ran, Feng - 686
Rao, C.S.P - 1066
Rao, C.S.P - 1032
Rao, C.S.P - 1419
Rashidi Komijan, A - 645
Razeghi, Yasamin - 60
Razeghi, Yasamin - 75
Razmi, Jafar - 1101
Razmi, Jafar - 1109
Reclick, Daniel - 336
Reclick, Daniel - 423
Refflinghaus, Robert - 1192
Refflinghaus, Robert - 1218
Richter, Matthias - 105
Riemma, S. - 1411
Ristof, Kleber D. - 1200
Robson, Antony - 452
Rocha, Carlos - 436
Rocha-Lona, Luis - 786
Rocha-Lona, Luis - 1234
Rogalski, Sven - 461
Rogalski, Sven - 987
Rong, Yiming - 6
Rossi, Cesare - 383
Rossi, Cesare - 391
Rothe, Jan - 777
Russell, David W. - 5
Russell, David W. - 328

S
Saenz de Argandoña, E - 477
Safaei Ghadikolaei, A - 565
Safavi, S. Mohsen - 265
Safavi, Mohsen - 850
Said, Ahmed Husam - 1243
Saliba, Michael A - 399
Saliba, Michael A - 407
Salman, Mustafa Ramzi - 743
Salminen, Kai - 964
Salminen, Kai - 1269
Sandanayake, Y. G. - 760
Santos, Hugo G. - 122
Santos, F. D. C. M. - 1139
Sarangapani, J. - 1519
Sarip, Shamsul - 1003
Sarker, Mosharraf - 282
Salwar, Mohammed - 858
Sarwar, Mohammed - 865
Savino, Sergio - 383
Savino, Sergio - 391
Sawassi, Mohamed - 1365
Saygin, Can - 1519
Schermer, Léo - 122
Schraven, Daan F.J. - 1073
Schwenk, Johann - 303
Shamah, Rania - 1498
Shaw, Alan - 1011
Shaw, Terry - 1084
Shelton, Paul - 816
Shen, Ling - 1481
Shokri, Alireza - 1208
Shokri, Alireza - 1375
Shokuhyar, S. - 509
Sice, Petia - 1527
Siefel, Jan - 987
Simas, Henrique - 436
Skalicky, J - 250
Skawinski, Piotr - 171
Smith, Paul - 1251
Sobis, Anna - 1302
Soriano-Meier, Horacio - 786
Soriano-Meier, Horacio - 1234
Sormaz, Dusan - 726
Sormaz, Dusan - 1146
Srchova, Marcela - 179
Srchová, Marcela - 257
Steiner, Frantisek - 925
Stevenson, Jessada - 218
Strano, Salvatore - 383
Strano, Salvatore - 391
Street, Graham - 282
Sudheer Kumar, B. V. - 1066
Sulong, Abu Bakar - 970
Swider, J. - 336
Szécsi, Tamas - 1277
Szendrei, Danny - 1403

T
Teich, Cornelia - 105
Teich, Tobias - 1403
Tennety, Chandy - 1146
Terpenny, Janis - 1163
Thompson, Chris - 1386
Tlaie, N. S. - 881
Tokio, T. - 1356
Tonetto, Cristiane P. - 130
Tupa, Jiri - 52
Turkyilmaz, Ali - 319

U
Unger, Katja - 1403
Uran, Miro - 939

V
Valente, Anna - 1356
Van Aken, Eileen - 1226
Van der Krogt, Roman - 743
Venkatadri, Uday - 581
Verl, Alexander - 226
Verl, Alexander - 1286
Vitiello, C. - 1356
Vosniakos, G.C. - 139
Vosniakos, G.C. - 1309

W
Wacker, Daniel - 461
Wagenknecht, C. - 1309
Wahab, D.A - 524
Wahab, D.A - 1124
Wahab, D.A. - 1155
Walther, Michael - 1286
Wang, Lihui - 186
Wang, Lihui - 360
Wang, Yi-Chi - 1491
Wang, W.H.A - 1458
Weigert, Gerald - 1325
Weissenberger-Eibl, M - 303
Williams, Stephen - 1386
Woods, Robert - 912
Worley, June - 1226
Wszolek, Grzegorz - 336
Wu, Yifan - 1476
Wyrick, David - 613
Wyrick, David - 800

X
Xu, Donglai - 694
Xu, Meihua - 680
Xu, Meihua - 686
Xu, Qiang - 919
Xu, Qiang - 932
Xu, Qiang - 939
Xu, Wendan - 694

Y
Yamamoto, Hidehiko - 524
Yamamoto, Yuji - 751
Yang, Gyun-Eui - 429
Yang, Gyun-Eui - 904
Yazdi, Saloomeh - 645
Yuan, Xumei - 1481

Z
Zaim, Halil - 319
Zaim, Selim - 319
Zain, R. M. - 469
Zammit, Dawn - 407
Zammit, Martin - 407
Zampognaro, Luca - 1333
Zandieh, M. - 493
Zandieh, M. - 501
Zarb, Andrew Vella - 399
Zarei, Sara - 351
Zarei, Sara - 452
Zareshahi, Ali - 565
Zhang, Jianyong - 636
Zhang, X. - 163
Zhou, Bin - 636
Zuriarrain, I - 477