



# *University of* **HUDDERSFIELD**

## **University of Huddersfield Repository**

Chen, Wen-hua, Chen, Xun and Xu, Zhijie

Proceedings of the 17th International Conference on Automation & Computing

### **Original Citation**

Chen, Wen-hua, Chen, Xun and Xu, Zhijie (2011) Proceedings of the 17th International Conference on Automation & Computing. University of Huddersfield. ISBN 9781862180987

This version is available at <http://eprints.hud.ac.uk/id/eprint/11542/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

# Proceedings of the 17<sup>th</sup> International Conference on Automation & Computing

---

University of Huddersfield, Huddersfield, UK, 10 September 2011

Edited by

*Wen-hua Chen*

*Xun Chen*

*Zhijie Xu*



---

The Chinese Automation and Computing Society in the UK

# Preface

Welcome to the 17th International Conference on Automation and Computing (ICAC'2011). The conference has a long successful history since its first conference in London (1995). The conference initially aims to provide a forum for Chinese scientists, engineers, scholars and students in the UK to update technical knowledge and exchange ideas, and to provide a platform where joint research programmes between those in both the UK and China can be formulated for mutual benefit. Since 2007, the conference had expanded towards a true international conference. This year, the conference brings together all researchers throughout the world, including Bhutan, China, Egypt, India, Iran, Romania, Switzerland and others, to the UK to disseminate their scientific findings in all aspects of automation, engineering, and computer science. We hope you have enjoyable time and find the conference stimulating and fruitful.

In general, the scope of the conference covers all the aspects of automation and computing from fundamental research to engineering applications and advanced technologies. As the tradition, we choose one research topic as the main theme of the conference each year which will highlight the important issue need to be addressed. This year, the committee has chosen “Computational Intelligence”. This is a core element for the engineering automation in the information era. We invite Professor Bernard Hon from the University of Liverpool to make a keynote speech on “Virtual Manufacturing” to present the opportunities and challenges of computational intelligence for industrial application. We also invite Professor Qingsheng Xie, Professor Ziqin Wang and Dr Weijie Pan to present us the “Opportunities and Challenges of Developing Equipment Manufacturing Industry in Western China”. The Chinese government has invested heavily to build a good industrial infrastructure and environment, to strengthen opening to the outside world, to attract foreign talents and technologies and to promote the equipment manufacturing industry development. We believe these keynote speeches will be very interesting.

After reviewing, 61 papers have been selected in the conference proceedings and accepted for the conference presentation. We are pleased to see the high quality of these papers, which contain both high level theoretical papers and practical application papers.

On behalf of the organising committee, we would like to take this opportunity to thank all those who have contributed to this conference, as well as the members of the organising committee and the international program committee for their fantastic work. Our sincere thanks also go to the conference sponsors for their suggestions and supports. Finally, but not least, we want to thank many volunteers for their diligent works for the conference.

Conference Chair: Dr Xun Chen

Programme Chair: Dr Wen-hua Chen

September 2011

# **Proceedings of the 17th International Conference on Automation & Computing (ICAC'2011)**

Held in University of Huddersfield, Huddersfield, UK. on 10 September 2011

**Organizer:** The Chinese Automation & Computing Society in the UK (CACSUK)

## **Conference Chair**

Dr. Xun Chen, University of Huddersfield

## **Programme Co-chairs**

Dr. Wen-hua Chen, Loughborough University

## **Students Activity Co-chairs**

Prof. Shuanghua Yang, Loughborough University

Prof. Jihong Wang, University of Warwick

## **Conference Co-Chair**

Dr Zhijie Xu, University of Huddersfield

## **Finance Co-Chairs**

Dr Lili Yang, Loughborough University

## **Local Committee**

Dr Fengshou Gu, University of Huddersfield

Dr Joan Lu, University of Huddersfield

## **Programme & Organizing Committees**

Yonghong Bu, Oxford University

Yi Cao, Cranfield University

Bo Chen, University of Warwick

Mou Chen, NUA

Sheng Chen, University of Southampton

Wen-Hua Chen, Loughborough University

Xun Chen, University of Huddersfield

Kai Cheng, Brunel University

Xiaochun Cheng, University of Reading

Yang Dai, University of Coventry

Xian Fei, University of Coventry

Jian Fu, Loughborough University

Dawei Gu, University of Leicester

Dongbing Gu, University of Essex

Huosheng Hu, University of Essex

Wei Huang, University of Bedfordshire

Yi Huang, University of Liverpool

Yanmei Huo, Jilin University

Yaochu Jin, University of Surrey

Ziqiang Lang, University of Sheffield

Dayou Li, University of Bedfordshire

Maozhen Li, Brunel University

Qingde Li, University of Hull

Yun Li, University of Glasgow

Guoping Liu, Glamorgan University

Honghai Liu, University of Portsmouth

Ji Yin Liu, Loughborough University

Julian Liu, Oxford University

Xiangdong Ma, University of Lancaster

Geyong Min, University of Bradford

Sheng Feng Qin, Brunel University

Gui Yun Tian, University of Newcastle

Bing Wang, University of Hull

Hong Wang, University of Manchester

Jin Wang, Liverpool John Moores University

Jihong Wang, University of Warwick

Wenbin Wang, University of Salford

Zidong Wang, Brunel University

Mianhong Wu, University of Derby

Qinghua Wu, University of Liverpool

Wenyan Wu, Staffordshire University

Dongling Xu, University of Manchester

Zhijie Xu, University of Huddersfield

Song Yan, University of Bedfordshire

Hongji Yang, De Montfort University

Jianbo Yang, University of Manchester

Lili Yang, Loughborough University

Shuang Hua Yang, Loughborough University

Taicheng Yang, University of Sussex

Hujun Yin, University of Manchester

Dingli Yu, Liverpool John Moores University

Hongnian Yu, Staffordshire University

Yong Yue, University of Bedfordshire

Jie Zhang, University of Newcastle

Sijing Zhang, University of Bedfordshire

Quanmin Zhu, University of West England

Ziqiang Zhu, University of Sheffield

Arygrios Zolotas, University of Loughborough



## **Advisory Committee**

Xing Ai (Shandong Univ., Jinan, China)	Xinping Yan (Wuhan Uni. of Technology, China)
Tianyou Chai (Northeast University, China)	Ya-Xiang Yuan (Inst. of Com. Maths. & Sci/.Eng)
Quangen Fang (Shanghai Maritime Univ., China)	Bo Zhang (Tsinghua University, Beijing, China)
Weihua Gui (Central South University, China)	Xiaoming Liu (Chinese Embassy, London)
Lei Guo (Chinese Academy of Sciences, China)	Hong Li (Chinese Embassy, London)
Guojie Li (CAS, Beijing, China)	Jian Ni (Chinese Consulate General, Manchester)
Zhongzhi Shi (CAS, Beijing, China)	Hengsheng Fu (Chinese Consulate General, Manchester)
Min Tan (Chinese Academy of Sciences, China)	Zhaosheng Wu (Chinese Consulate General, Manchester)
Tieniu Tan (Chinese Academy of Sciences, China)	
Youlun Xiong (Huazhong Univ. of Sci. & Tech.)	

## **Sponsors:**

The Education Section of Chinese Embassy in London,  
The Education Section of Chinese Consulate General in Manchester,  
The University of Huddersfield,  
Sinocera Piezotronics Inc.,  
Beijing Zhongguancun Science Park

# Table of Contents

## Key notes:

*Opportunities and Challenges of Developing Equipment Manufacturing Industry in Western China, .....xiii*

*Professor Ziqin Wang, Guizhou University, Guizhou Province, PR China*

*Virtual Manufacturing – Current Advances and Industrial Applications .....xiv*

*Professor Bernard Hon, University of Liverpool, UK*

## Theme 1: Computing and Network

*An Intelligent Cloud Computing Architecture Supporting e-Governance ..... 1*

*Rajkumar Sharma, Vikram University, India*

*Priyesh Kanungo, Patel College of Sc. & Technology, India*

*Toward Optimal Multi-Objective Models of Network Security: Survey ..... 6*

*Valentina Viduto, Wei Huang, Carsten Maple, University of Bedfordshire,UK*

*A Complete Planner Design of Microstrip Patch Antenna for a Passive UHF RFID Tag ..... 12*

*Tashi, Royal University of Bhutan, Rinchending, Phuentsholing, Bhutan & Staffordshire University, UK*

*Mohammad S. Hasan, Hongnain Yu, Staffordshire University, Becondside, UK*

*A fast and effective way to improve the merging accuracy of multi-view point cloud data ..... 18*

*Feng Li, Andrew Longstaff, Simon Fletcher, Alan Myers, Huddersfield University, UK*

*Representing the Process of Machine Tool Calibration in First-order Logic ..... 22*

*S. Parkinson, A .P. Longstaff, A. Crampton, S. Fletcher, G. Allen, A. Myers, University of Huddersfield, UK*

*Localisation Algorithm in Wireless Sensor Networks ..... 28*

*Shuang Gu, Yong Yue, Carsten Maple, University of Bedfordshire, UK*

*Chengdong Wu, Northeastern University, China*

*User-defined gesture sets using a mobile device for people with communication difficulties ..... 34*

*Yong Hee Jung, Shengeng Qin, Brunel University, , UK*

The Effect of Depth of Cut on the Molecular Dynamics (MD) Simulation of Multi-Pass Nanometric Machining..... 40

*A.O. Oluwajobi and X. Chen, University of Huddersfield, UK*

Mobile Motion Gesture Design for Deaf People ..... 46

*Haoyun Xue, Shengfeng Qin, Brunel University, UK*

An Efficient and Secure Authentication Protocol for RFID Systems ..... 51

*Md. Monzur Morshed, Anthony Atkins, Hongnian Yu, Staffordshire University, UK*

Design of Interference Aware ZigBee Building Monitoring Network..... 57

*Fang Yao and Shuang-HuaYang, Loughborough University, UK*

Species area relations and information rich modelling of plant species variation ..... 63

*James Furze, Quan Min Zhu, Jennifer Hill, University of the West of England, UK*

*Feng Qiao, Shenyang Jianzhu University, China*

## **Theme 2: Advanced Control & Applications**

Study of a Multivariable Coordinate Control for a Supercritical Power Plant Process ..... 69

*Omar Mohamed, University of Birmingham, UK*

*Jihong Wang, University of Warwick, UK*

*Bushra Al-Duri, University of Birmingham, UK*

Robust Stability and performance with  $H_2/H_\infty/\mu$ controller for Single Person Aircraft ..... 75

*J. Mashayekhi Fard, Islamic Azad University, Iran*

*M.A. Nekoui, A. Khaki Sedigh, K. N. Toosi University of Technology, Iran*

*R. Amjadifard, Tarbiat Moallem University, Iran*

Grid Roadmap based Real time Path Planning ..... 81

*M. R. B. Bahar, H. B. Bahar & F. Hashemzadeh, Tabriz University, Iran*

A New Method for Estimating the Maximum Allowable Delay in Networked Control of Bounded

Nonlinear Systems.....	86
<i>Ashraf F. Khalil, University of Birmingham, UK</i>	
<i>Jihong Wang, University of Warwick, Coventry, UK</i>	
Saliency Investigation of PM Brushless AC Motors for High-Frequency Carrier Signal Injection-Based Sensorless Control .....	92
<i>Liming Gong and Z. Q. Zhu, University of Sheffield, UK</i>	
$H_\infty$ filter design of networked systems with uncertain accessing probabilities and quantisation.....	98
<i>Hongbo Song, Li Yu, Zhejiang University of Technology, China</i>	
<i>Guo-ping Liu, University of Glamorgan, UK</i>	
Sliding Mode Control for a miniature helicopter.....	104
<i>Jian Fu, Qing-xian Wu, Nanjing University of Aeronautics and Astronautics, China</i>	
<i>Wen-hua Chen, Loughborough University, UK</i>	
Real-Time Implementation of a Burst Error Compensator for Wireless Control Systems .....	110
<i>Michael Short, Usama Abrar, Ian French and Fathi Abugchem, Teesside University, UK</i>	
Neural Generalized Predictive Controller and Internal Model Principle.....	116
<i>Hesham Abdel-Ghaffar, Invensys Engineering &amp; Service, Egypt</i>	
<i>Sherif Hammad, Hazem Abbas, Mentor Graphics Corporate, Egypt</i>	
<i>A.Z. Badr, Ahmed Hassan, Ain Shams University, Egypt</i>	
State Observer in Networked Control Systems with Variable Delay in the Feedback Channel .....	122
<i>A. Sedighi, R. Mahboobi Esfanjani, Sahand University of Technology, Iran</i>	
Stabilization of Networked Control Systems with Variable Transmission Delays .....	126
<i>M. Mahmodi Kaleybar, R. Mahboobi Esfanjani, Sahand University of Technology, Iran</i>	
Zero Overshoot and Fast Transient Response Using a Fuzzy Logic Controller .....	130
<i>Bakhtiar I. Saeed &amp; Bruce Mehrdadi, University of Huddersfield, UK</i>	

Trajectory Generation for Autonomous Soaring UAS.....	135
---	-----

*J. H. A. Clarke and W-H. Chen, Loughborough University, UK*

Multi-objective Optimization of Constrained Parallel Hybrid Electric Vehicles.....	141
--	-----

*Shaobo Li, Jinglei Qu, Ministry of Education, Guiyang, China*

*Guanci Yang, Chinese Academy of Sciences, Chengdu, China*

### **Theme 3: Intelligent Manufacturing**

Morphological Filters Based on Motif Combination for Functional Surface Evaluation .....	147
--	-----

*Shan Lou, Xiangqian Jiang, Paul J. Scott, University of Huddersfield, UK*

Data Mining for Gearbox Condition Monitoring .....	152
--	-----

*M. Baqqar, M.Ahmed , F. Gu, The University of Huddersfield, UK*

Process Monitoring and Metrology for Single Grit Grinding Test Performance .....	157
--	-----

*Tahsin Tecelli Öp öz and Xun Chen, University of Huddersfield, UK*

Improving Control Panel Consistency of Wizard-of-Oz Design and Evaluation Studies .....	163
---	-----

*Andol X. LI and John V. H. BONNER, University of Huddersfield, UK*

Assessing customized product design using virtual human and imposed motion .....	169
--	-----

*Shengfeng Qin, George Panayiotou, Brunel University, UK*

*Pin Zhang, Nankai University, China*

Volume Deformation Based on Model-Fitting Surface Extraction .....	175
--	-----

*Qian Xu, Duke Gledhill, Zhijie Xu, University of Huddersfield, UK*

Numerical Simulation of Natural Frequencies in the Design of Micro Air Vehicle structures .....	181
---	-----

*Yanan Yu, Xiangjun Wang, Tianjin University, China*

*Carlo Ferri, Qingping Yang, Brunel University, UK*

Finite Element Investigation of Nano-indentation of coated Stainless Steel .....	186
--	-----

*Qiang Xu, Chulin Jiang, Dezheng Liu, Yongxin Pang, Simon Hodgson, Teesside University, UK*

The development and Validation of Multi-axial Creep Damage Constitutive Equations for P91 .....	191
---	-----

*Qiang Xu, Mark Wright, Qihua Xu, Teesside University, UK*

#### **Theme 4: Computational intelligence**

Short-Term Load Forecasting System Using Data Mining .....	197
--	-----

*Liu Jin, Yu Jilai, Harbin Institute of Technology, China*

Applying the Design of Experiment (DoF) to Optimise the NN Architecture in the Car Body Design System .....	203
---	-----

*Sugiono, Mian Hong Wu, Ilias Oraifige, University of Derby, UK*

Multiobjective Design of Evolutionary Hybrid Neural Networks .....	209
--	-----

*Lavinia Ferariu and Bogdan Burlacu, "Gheorghe Asachi" Technical University of Iasi, Romania*

A regressive schema theory based tool for GP evolved nonlinear models .....	215
---	-----

*Alina Patelli, Lavinia Ferariu, "Gheorghe Asachi" Technical University of Iasi, Romania*

Vehicle Windscreen Wiper Mathematical Model Development and Optimisation for Model Based Hardware-in-the-Loop Simulation and Control .....	221
--	-----

*Jianlin Wei, Jihong Wang, Hao Sun, University of Warwick, UK*

*Alexandros Mouzakitis, University of Warwick, UK*

Fault Classification of Reciprocating Compressor Based on Neural Networks and Support Vector Machines .....	227
---	-----

*M. Ahmed, S. Abdusslam, M. Baqqar, F. Gu, A.D. Ball, University of Huddersfield, UK*

Reinforcement Learning based Radio Resource Scheduling in LTE-Advanced .....	233
--	-----

*Ioan S. Comşa, Mehmet Aydin, Sijing Zhang, University of Bedfordshire, UK*

*Pierre Kuonen, Jean-Frédéric Wagen, University of Applied Sciences of Western Switzerland,*

Switzerland

## Theme 5: Advanced Sensor & Measurement Techniques

The extraction of characteristic quantity of shallow defects in pulsed magnetic flux leakage signal . 239

*Junbiao Fei, Xianzhang Zuo, Tao Zhang, Ordnance Engineering College, China*

*Yunze He, Guiyun Tian, Newcastle University, UK*

Detection technology to identify money based on pulsed eddy current technique ..... 244

*Sumin Qian, Xianzhang Zuo, Ordnance Engineering College, China*

*Yunze He, Guiyun Tian, Hong Zhang, Newcastle University, UK*

Infrared Thermography Study of Thermal Plume ..... 248

*Jafar Ali, Abdullah Abuhabaya and John Fieldhouse, University of Huddersfield, UK*

A Wireless Sensor Network based Structural Health Monitoring System for an Airplane ..... 254

*Jasleen. K. Notay, Ghazanfar. A. Safdar, University of Bedfordshire, UK*

Modelling and experimental investigation of ferromagnetic material for angular defect detection ... 260

*Dong Chang, Xianzhang Zuo, Ordnance Engineering College, China*

*University of Newcastle upon Tyne, UK*

Defect depth effects in Pulsed Eddy Current thermography ..... 265

*Hong Zhang, Guiyun Tian, Yunze He, Newcastle upon, UK*

*Xianzhang Zuo, Ordnance Engineering College, Shijiazhuang, China*

Parameters Influence in Steel Corrosion Evaluation Using PEC Thermography ..... 269

*Yunze He, Guiyun Tian, Liang Cheng, Hong Zhang, Newcastle University, UK*

*Paul Jackson, International Paint Ltd. Gateshead, Tyne and Wear, UK*

The Effect of Metallic Substance on the Read Range of UHF Passive RFID System ..... 275

*Chencho, Royal University of Bhutan, Bhutan*



*Chencho, Dr. Justin Champion, and Prof. Hongnian Yu, Staffordshire University,UK*

Electromagnetic NDT and Condition Monitoring – A Personal View ..... 280

*Xiandong Ma, Lancaster University,UK*

## **Theme 6: Intelligent Systems**

A Constraint-based Design Risk Management Tool for Design Collaboration ..... 286

*Jian Ruan, Sheng Feng Qin, Brunel University,UK*

Mixed integer linear programming models for scheduling the LED planting operation on PCBs ..... 291

*Jiaxiang Luo, South China University of Technology, China*

*Jiaxiang Luo, Jiyin Liu, Loughborough University, UK*

A NOVEL APPROACH TO MODELLING AND SIMULATION OF THE DYNAMIC BEHAVIOUR  
OF THE WHEEL-RAIL INTERFACE ..... 297

*Arthur Anyakwo, Crinela Pislaru, Andrew Ball, Fengshou Gu, University of Huddersfield, UK*

Time Encoded Signal Processing and Recognition of Incipient Bearing Faults ..... 303

*S. Abdusslam, M. Ahmed, P. Raharjo, F. Gu, A. D. Ball, University of Huddersfield, UK*

Reviewing DSTATCOM for Smart Distribution Grid Applications in Solving Power Quality Problems  
..... 308

*Bala Boyi Bukata and Yun Li, University of Glasgow,UK*

Analysis of interrelationships between Suppliers configuration and performance within the supply  
network: A simulation approach ..... 314

*Maria Aina, Dr. Yang Dai, Professor Dobrila Petrovic, Coventry University,UK*

## **Theme 7: Energy Conversion and Powertrain**

Review Electromagnetic Field Ignition System for Internal Combustion Engine ..... 320

*Lutz-Christoph Schöning and Yun Li, University of Glasgow,UK*

The Optimisation of Bio-diesel Production from Sunflower Oil using RSM and its Effect on Engine Performance and Emissions.....	324
--	-----

*Abdullah Abuhabaya, Jafar Ali, John Fieldhouse, Rob Brown and Eko Andrijanto, University of Huddersfield, UK*

Homogeneous Charge Compression Ignition engine: A Technical Review.....	329
---	-----

*Hammad Iqbal Sherazi and Yun Li, University of Glasgow, UK*

Energy Management System for Tribid Electric Vehicles.....	335
--	-----

*Kary Thanapalan, Fan Zhang , University of Glamorgan, UK*

Keynote Speeches:

# Opportunities and Challenges of Developing Equipment Manufacturing Industry in Western China

Qingsheng Xie, Ziqin Wang and Weijie Pan  
Guizhou University, China

**Abstract:** The Speech briefly introduces the overall development situation of Chinese equipment manufacturing industry in facing a new round of the great western strategic development and the favourable opportunities in the domestic and international industrial structural adjustment and transfer, which brings opportunities and challenges to the equipment manufacturing industry development in western China. According to the strategic position of the existing basis and conditions of the equipment manufacturing industry in Guizhou, western China, the paper explores the roadmap to develop the equipment manufacturing industry in Guizhou, and suggest the way to build a good environment, to optimize industrial layout and structure, to strengthen opening to the outside world, to attract foreign talents and technologies and to promote the equipment manufacturing industry development better and faster in Guizhou.

**Professor & Dr QingSheng Xie** was born in October, 1954, at Guiyang, Guizhou, PR China. He is present vice-governor of the People's Government of Guizhou Province, PR China. Professor Xie QingSheng is a young and middle-aged expert who has outstanding national contributions. He mainly engaged in teaching, research and development of manufacture information system, served as a national 863 expert group expert and participated in many national and local information strategic project researches. He published more than 100 papers and 3 books. He has successively presided more than 40 significant scientific research projects.

**Professor & Dr ZiQin Wang** was born in 1954, who is a Ph.D. supervisor. He is a deputy director of the Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, a deputy director the Key Laboratory of the Advanced Manufacturing Technology Guizhou Province, and a director of Laboratory Centre of Mechanical Engineering College. He has visited the United States twice as a visiting scholar. His main research field covers manufacturing technology and equipment, CAE technology. He has two provincial science and technology progress prizes, five national patents, and more than 40 published papers.

Keynote Speeches:

# Virtual Manufacturing – Current Advances and Industrial Applications

Bernard Hon  
University of Liverpool  
Liverpool, UK

**Abstract:** It is now more than 40 years since the pioneering work on virtual reality (VR) was launched. The technology has progressed to new level of technical competence and it is now part of the new product development process for world class manufacturers in the automotive and aerospace industry. This keynote will outline the state-of-the-art developments in VR followed by a concise summary on the distinctive features of VR. It will also cover recent developments in augmented reality and its emerging usage. A whole range of activities of the new multi-million pound Virtual Engineering Centre (VEC) will be reported. A detailed case study on the use of DELMIA for virtual assembly modelling and analysis of an air-blower at the VEC is given. The virtual assembly approach allows detailed study on energy expenditure, posture analysis and cycle time for workplace design and assembly process optimization. This virtual manufacturing approach will form part of the pre-production planning for productivity improvement and cost reduction.

**Professor Bernard Hon** is currently Professor of Manufacturing Systems at the University of Liverpool in the UK. His research is focused on advanced manufacturing technology, sustainable manufacturing and manufacturing systems analysis. Prof. Hon has served on numerous professional institutions, national and international research organizations. He also acts as consultants to industry on new product and process development. He is a Fellow of the International Academy for Production Engineering (CIRP) and the Institution of Engineering and Technology.

# An Intelligent Cloud Computing Architecture Supporting e-Governance

Rajkumar Sharma<sup>1</sup> and Priyesh Kanungo<sup>2</sup>

1. Vikram University, Ujjain, India  
rksujn@rediffmail.com

2. Patel College of Sc. & Technology, Indore, India

**Abstract**—The development of high speed Internet access, Web 2.0 applications and Virtualization techniques have made Cloud computing a leading edge technology. A user in ‘Cloud’ runs web based application over Internet via browser with a look and feel of desktop program. Cloud computing provides dynamically scalable and virtualized resources as a service over the network at a nominal initial investment. Data-center works as backbone in Cloud computing where a large number of servers are networked to host computing & storage needs of the users. The area which needs more attention is Latency Optimization for cloud architecture to work as ubiquitous as expected. Many data intensive applications produce enormous amounts of data which travel on cloud network. As the cloud users grow, cloud architecture should accommodate movement of voluminous data to avoid data congestion in the network. In this paper, an intelligent & energy efficient Cloud computing architecture is proposed based on distributed data-centers to support application and data access from local data-center with minimum latencies. It was found that the proposed architecture is efficient for business entrepreneurs, suitable to apply for e-Governance and provides a green eco-friendly environment for Cloud computing.

**Keywords**—cloud computing; green computing; latency

## I. INTRODUCTION

Computer scientists have always been attempting and innovating a new technology that efficiently & effectively utilizes the contemporary underlying hardware resources for the benefit of the science and business community. Starting from mainframes to recent virtual machines on ‘Clouds’, computational history experienced a trend of alternatively convergent and divergent patterns for the use of computing resources. Mainframe/Mini Computers processed users programs centrally on time sharing concept. The deep penetration of cheap Personal Computers affected almost every corner of computing thus diverging the resources. Later, by again converging the computing resources as shown in Fig. 1, dedicated parallel machines run parallel programs faster than contemporary PCs. As these parallel machines were very expensive, a major computational change was observed as

divergence in resources which categorized as distributed computing such as Network of Workstations (NOW), Cluster computing, Grid Computing etc. In present scenario, Cloud computing uses centralized resources in form of data-center. The trend in use of resources alternatively, i.e., centralized and distributed seems to continue.

### A. Cloud Computing

Cloud computing is use of scalable computing resources over Internet on a pay-as-you-go basis [1]. It provides a cost-effective IT solution to business & scientific community. Economically the main attraction from Cloud computing is that customers only use what they need, and pay for what they actually use. Organisations neither need to purchase expensive hardware such as servers, storage, networking equipments etc. nor require manpower for development of complex IT solutions in-house. Resources as a service are available over Cloud at any time and from any location via the Internet [9]. Passive ‘Consumers’ have now become ‘Prosumers’ by using Service Oriented Architecture (SOA) and Web 2.0 applications such as social networking sites, blogs, hosting services etc. One can choose services from pool of available services and negotiate price through Service Level Agreements (SLAs). Among the popular Cloud service providers are: Amazon [5], Google [6], Microsoft [7] etc.

Three main types of service levels as delivery models are:

1) *Software as a Service (SaaS)* : The clients may opt for ready customized application, but do not have control over background environment such as operating system, hardware or network parameters.

2) *Platform as a Service (PaaS)* : In this types of services, clients have control over change in application and hosting environment such as system software. But SaaS does not provide control over operating system, hardware and network parameters.

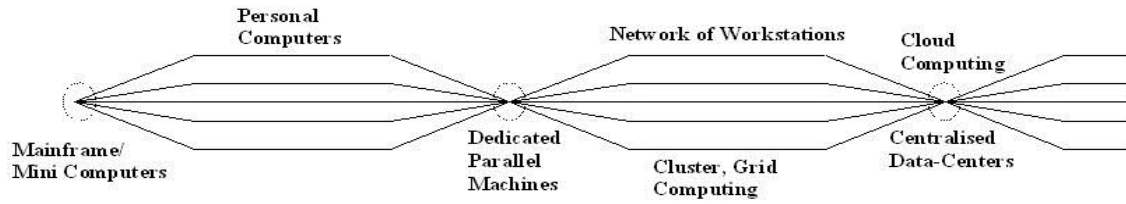


Figure 1. Convergent and Divergent Trends in Computing History

3) *Infrastructure as a Service (IaaS)* : The clients can create a virtual processing environment by specifying choice of processing power, storage, network parameter etc. and have control over operating system and application environment.

### B. Virtualization Technology

With virtualization techniques, multiple operating systems can concurrently run on a single physical system. A user can opt for his choice of operating system and other hardware configuration called virtual machine (VM) and run his application by sharing underlying hardware resources. It is the 'Virtual Infrastructure Management Software' (VIMS) that centrally manages many VMs on a single physical system. In user's perception each VM is a single, logical bunch of resources as shown in Fig. 2. Virtualization techniques provide cost-effective & efficient utilisation of IT infrastructure. Presently, Xen (<http://www.xen.org>) [10] and VMWare (<http://www.vmware.com>) [11] are two leading virtualization technology providers.

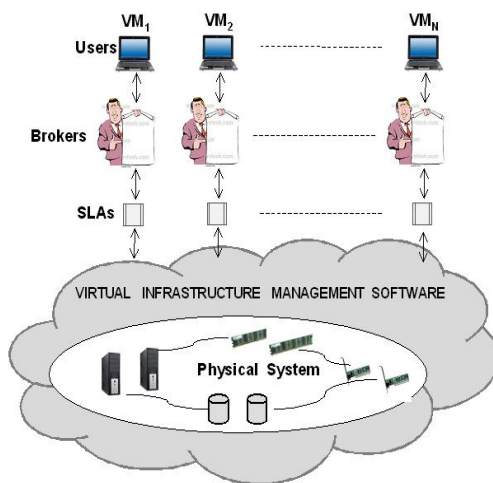


Figure 2. Virtual Machines in Cloud Architecture

### C. e-Governance

e-Governance is use of IT infrastructure to ease governance activity such as administration, revenue services, various services to citizens, policy formation etc. e-Governance improves the efficiency of government functioning by removing redundancy at different levels. Citizens get advantage from several e-services like income tax, pension, services related to municipal corporation and agriculture etc. The four main categories of e-Governance applications and the services fall under these categories are :

- Government to Government (G2G): Administration, Policy formation etc.
- Government to Business (G2B): Taxation, Tender etc.
- Government to Consumer (G2C): Land record, Birth cetificate etc.
- Government to Employees (G2E): Income tax, Pension etc.

Transforming ongoing national e-Governance plan of governments to Cloud architecture would yield the following benefits :

- Uniform e-Governance architecture all over the country as against the heterogeneous architecture due to procurement of IT infrastructure from autonomous government agencies.
- Governments need not have expensive IT setups, pay for software licenses and maintain them leading to a substantial cut in the government's annual budget for IT infrastructure.
- Governments can concentrate on making e-Governance convenient for the intended users instead of mere focus on technical and operational overheads to maintain IT infrastructure.

## II. RELATED WORK

Continuous research work is being carried out worldwide to address the issue of Cloud computing architecture. Some researchers identify Cloud computing as virtualization of previously existing data-centers while some others nominate data-centers as backend resources of newly adopted Cloud computing paradigm. Fang & Yin [2] correlate Cloud architecture with Business Process Management (BPM). Business process activities are modeled over platform layer and combined with application layer. A strategy to manage complex and unpredictable workload entering cloud is proposed by Paton et al.[3]. As all computing & storage resources are managed centrally, despite workload balance, the system is susceptible to network congestion. Dabas & Gupta [4] propose a Cloud architecture for radio frequency identification. The data captured by radio frequency reader is sent to data processing system present in the cloud. A substantial time delay may be observed if radio frequency reader and cloud resources are physically located at long distance. A common trend of centralized resources at the Cloud provider's location is present in almost all existing Cloud computing architectures leading to increase in latencies.

## III. PROPOSED CLOUD ARCHITECTURE

Due to world-wide hype and rapid growth in associated technologies, Cloud computing clients continue to multiply. The large number of service requests to fulfill the demands of millions of users will broaden the latency problem. Cloud service provider physically may be far away from the clients, compelling

data to travel from several mediums and network equipments, thereby imposing a time delay in getting Cloud services. Existing Cloud providers use centralized data-center to host computing & storage needs of the clients. In this study, an intelligent & energy efficient Cloud computing architecture is proposed based on distributed data-centers which form a client's instance in nearest neighborhood and fulfill client's request in optimized latency.

### A. Cloud Computing Model

In the proposed Cloud architecture data-centers work in master-slave paradigm. Nearest data-centers form a computing zone and users may opt for creating their instances in multiple zones. The main entities involved in proposed architecture are :

- 1) *Master/Slave Data-Center*: Master data center is located at Cloud provider's administrative premises. User's accounting on pay-as-you-go basis is completed here. Slave data-center are geographically scattered to serve user's requests in minimum physical distance.
- 2) *Users/Brokers* : Users directly communicate or via brokers submit requests which automatically reaches at master data-center. Master data-center creates user instance at appropriate slave data-center considering minimum latency.
- 3) *Service Level Agreements (SLAs)* : Quality of Service (QoS) and pricing negotiations are settled through SLAs. Master data-center scans SLA each time to host needs of the users.

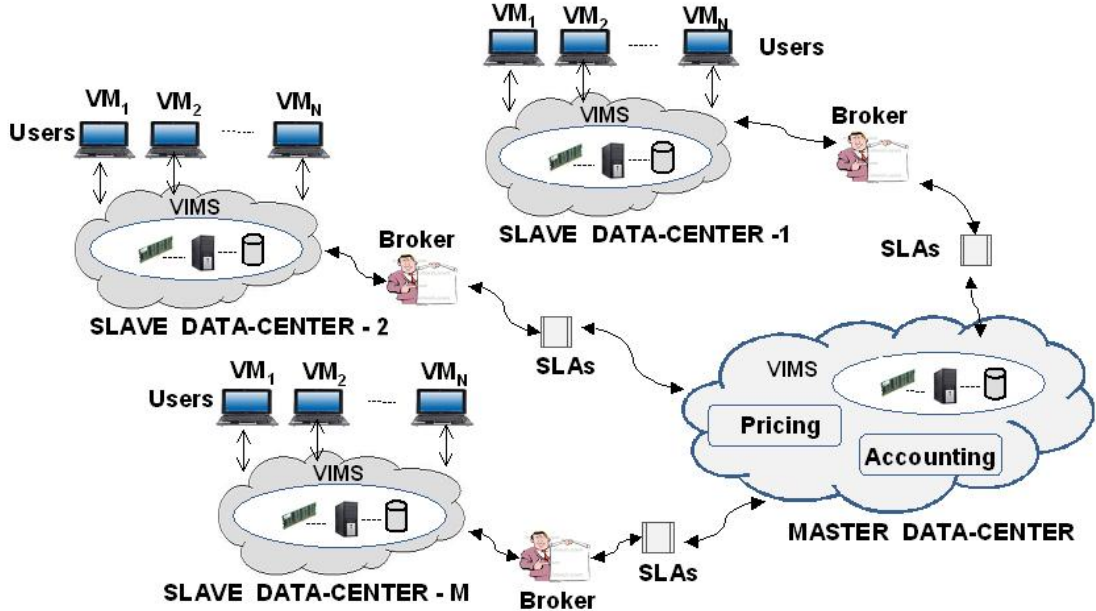


Figure 3. Proposed Cloud Architecture



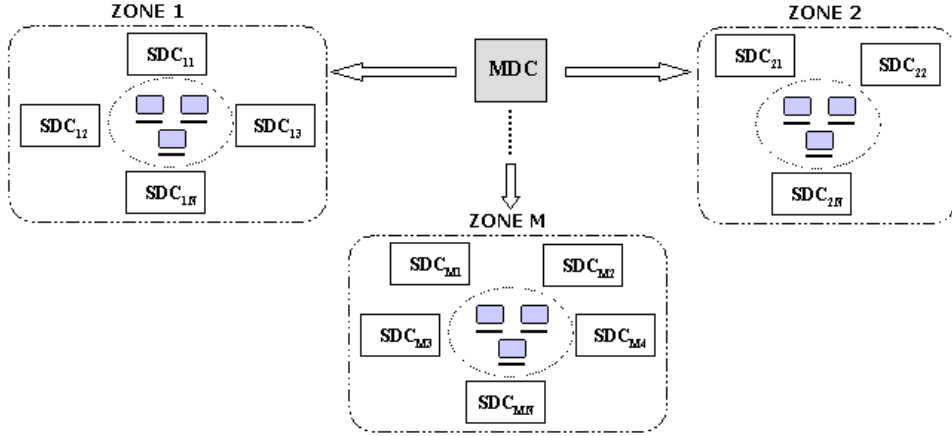


Figure 4. Creation of Instance Based on User's Geographic Location

### B. Informal Description of Algorithm

After receiving a request for creating an instance, master data-center look for the availability of resources in the user's local data-center. If desired resources are available, then user gets his required instance and run his application with minimum latency. Master data-center searches other slave data-centers of same zone for resources if they are not available on the location of the user. If resources are not available even in same zone and user has opted for multiple zones then master data-center looks for resources in other zones.

### C. Formal Description of Algorithm

If we denote  $SDC_{ij}$  as slave data-center in  $i^{th}$  zone and at  $j^{th}$  location ( $i=1..M, j=1..N$ ), -center, then the algorithm in proposed Cloud architecture for allocating instances to user  $MDC$  as master data-center and  $U_{pq}$  is an user in  $p^{th}$  zone and at the location of  $q^{th}$  slave data is :

Algorithm Create\_Instance

Request for instance from user  $U_{pq} \rightarrow MDC$

$MDC \rightarrow$  if resources available in  $SDC_{ij}$

$(i = p, j = q)$

create instance ( $U_{ij}$ ) ;

else if resources available in  $SDC_{ij}$

$(i = p, j = 1..N, j \neq q)$

create instance ( $U_{ij}$ ) ;

else if multiple zones = 'yes' search for resources

in  $SDC_{ij} (i = 1..M, i \neq p, j = 1..N)$

if resources available, create instance ( $U_{ij}$ );

else make a fresh request  $U_{pq} \rightarrow MDC$  ;

End of algorithm.

### D. Advantages of Distributed Data-Centers in Cloud

- Users get quick response to their requests in minimum latency
- Data-centers may be formed with commodity hardware as against expensive hardware in centralized data-center.
- Entrepreneurs may form local data-center on their existing IT infrastructure for security measures, thus forming a hybrid Cloud.
- A large number of services and network equipments are clustered in centralized data-center requiring more electricity, air conditioning etc., whereas distributed data-center require less power consumption, air conditioning etc., therefore creating an eco-friendly green computing environment.

### E. Suitability of Distributed Data-centers in e-Governance

- Distributed data-center may prove to be very effective in e-Governance as nature of government functioning is distributed
- Master data-center may be formed in capital of a state and slave data-center may be formed either at division or district places.
- Users may get a direct connectivity through fiber optical cable to local data-center thereby providing a maximum bandwidth.
- Governments may host their own Clouds, making local data-centers on the existing IT infrastructure.
- Government may increase revenue by hosting services on own Clouds and enforcing tax deduction at source in all business transactions.

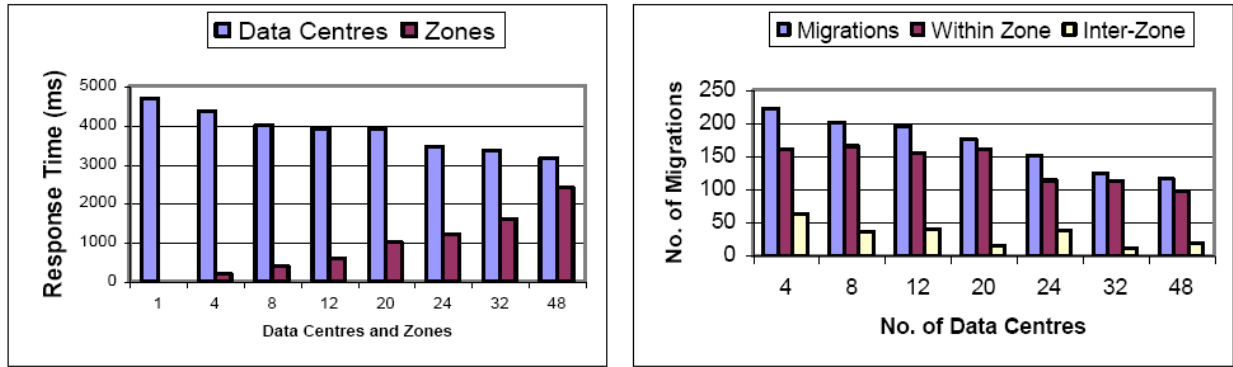


Figure 5. Comparison of Response Time and No. of Migrations.

#### IV. EXPERIMENTAL RESULTS AND COMPARISON

Experiments were carried out of the proposed architecture on Cloud simulator [8] which has provision for forming different data centers, virtual machine instance migration, energy consumption model etc. By running an application initially from central data center and then from several geographic locations on different number of data centers, about 21% improvement was observed in latencies as shown in Fig. 5. User's instances created as virtual machines are migrated across physical servers of data center for load balancing purposes. Less number of virtual machine migrations were reported as compare to central data center as computing resources are distributed across geographic locations. Maximum migrations occurred within zone of user's locality than inter-zone migrations. Barker & Shenoy [12] show data of read/write experiments for a latency sensitive multimedia application running on a single data center. We simulated a multimedia application running on platform consists of distributed data centers which outperforms about 12% the results of [12] in terms of response time similar to as shown in Fig. 5.

Data center usually consists of computing, storage & network devices almost in thousands of numbers. In addition to huge amount of electricity consumed by these devices, huge electric power is also needed for cooling systems. As a result data center emits more carbon dioxide (CO<sub>2</sub>) in a locality than permissible emission standard of the government. Cloud architectures proposed by several authors including [3] and [4] are based on central data center. On the other hand, in the proposed distributed data centers carbon emission would be within permissible limit as the resources are scattered in wide geographical area.

#### V. CONCLUSION

Scientific and commercial computing have seen several architectural combinations of available contemporary computing resources. In the similar way as Grid computing gained much attention from scientific community, Cloud computing is being

popular among business community. The architectural design of Cloud computing is still in its infancy and needs exploration towards the efficient utilization of large scale IT infrastructure. In this paper, an effective Cloud computing architecture is presented based on distributed data-centers which yields access to application & data among users in minimum latencies and create an energy efficient & eco-friendly green computing environment. A possibility of using existing IT setups as local data-center make proposed architecture cost-effective for organizations.

#### REFERENCES

- [1] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities," *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC 2008)*, Dalian, China, Sept. 25-27, 2008.
- [2] Zhenyu Fang and Changqing Yin, "BPM Architecture Design Based on Cloud Computing," *Online Journal on Intelligent Information Management*, Vol 2, May 2010, pp 329-333.
- [3] Norman W. Paton, Marcelo de Aragao, Kevin Lee, Alvaro Fernandes and Rizos Sakellariou, "Optimizing Utility in Cloud Computing through Autonomic Workload Execution," retrieved from <http://research.microsoft.com/pub/debull/A09mar>.
- [4] Chetna Dabas and J.P Gupta, "A Cloud Computing Architecture Framework for Scalable RFID," *Proceeding of the International Multi Conference of Engineers and Computer Scientists (IMECS 2010)*, Hong Kong, Vol 1, March 2010, pp 217-220.
- [5] Amazon Elastic Compute Cloud (EC2), <http://www.amazon.com/ec2/>
- [6] Google App Engine, <http://appengine.google.com/>
- [7] Microsoft Live Mesh, <http://www.mesh.com/>
- [8] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose, and Rajkumar Buyya, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, Volume 41, Number 1, Pages: 23-50, New York, USA, January, 2011.
- [9] Suraj Pandey, "Cloud Computing Technology & GIS Applications," *Asian Symposium on Geographic Information Systems From Computer & Engineering View (ASGIS 2010)*, China, April 2010.
- [10] Virtualization Technology, <http://www.xen.com/>
- [11] Virtual Machines through VMware, <http://www.vmware.com/>
- [12] Sean K Barker, Prashant Shenoy, "Empirical Evaluation of Latency-sensitive Application Performance in the Cloud,"

# Toward Optimal Multi-Objective Models of Network Security: Survey

Valentina Viduto(Presenting Author), Wei Huang and Carsten Maple

Institute for Research in Applicable Computing (IRAC)

University of Bedfordshire

Luton, United Kingdom

[valentina.viduto@beds.ac.uk](mailto:valentina.viduto@beds.ac.uk), [wei.huang@beds.ac.uk](mailto:wei.huang@beds.ac.uk) and [carsten.maple@beds.ac.uk](mailto:carsten.maple@beds.ac.uk)

**Abstract** — Information security is an important aspect of a successful business today. However, financial difficulties and budget cuts create a problem of selecting appropriate security measures and keeping networked systems up and running. Economic models proposed in the literature do not address the challenging problem of security countermeasure selection. We have made a classification of security models, which can be used to harden a system in a cost effective manner based on the methodologies used. In addition, we have specified the challenges of the simplified risk assessment approaches used in the economic models and have made recommendations how the challenges can be addressed in order to support decision makers.

**Keywords**—*risk assessment; multi-objective models; network security optimisation; visualisation techniques; survey;*

## I. INTRODUCTION

Today's IT infrastructures, systems and applications are more integrated, dynamic and distributed. According to [1] we have reached the era where information is the key for business to thrive and never before, information has been so important. With the rapid development of Information Technologies (IT) and increased popularity to run business online, networked systems are becoming increasingly vulnerable to cyber attacks. As a result, attacks can influence the productivity, revenue and reputation. Thus, there is a need to know the possible chances of securing informational assets in the environment where the number of attacks and threats emerge rapidly [2]. The models developed recently are oriented towards better analysis and assessment of threats, risks, vulnerabilities and network security measures. But even with the presence of such models, security measures cannot always protect assets from threats. This is because of a poor network assessment and inherent management weaknesses. Hence, the security risk can never be fully eliminated, because it cannot be predicted. Though, the security models today should be designed to help security administrators and decision makers to take effective decisions when a number of constraints is considered.

Network risk assessment plays a crucial role in modern society and is one of the important processes of information security management. A risk management process is needed in order to identify, describe and analyse network vulnerabilities. The final goal of a standard risk

assessment procedure is to make security specialists and managers aware of the possible risks, network state and to help them to adopt security measures effectively. However, to be cost-effective, a coherent, well structured and straight forward risk assessment procedure should include the relationships among vulnerabilities, threats and countermeasures. The common methodologies are ISO 27001, ISO 17799 and guidelines issues by the National Institute of Standards and Technology NIST SP800-30 [3, 4, 5].

Usually two approaches can be used to assess systems for risks. One of them is a qualitative risk assessment approach. It uses modelling, visualization techniques to give an overview of the state a system holds, whilst quantitative approach tries to give a measure for risk. In recent literature, researches argue, that combining both approaches can be more beneficial than using these approaches separately [6].

With the high increase of automated systems and tools, humans have gained an ability to visualise, design and model networked systems, their security states, connection paths and other security related factors. Modelling techniques have been receiving a great interest between researchers. In terms of modelling, attack trees, attack graphs and other visualization techniques, such as onion skin model, offer a goal-oriented perspective of multi-stage attacks, help to estimate with the attack related costs, visualize dependencies between vulnerabilities and security measures [6, 7, 8, 9, 10]. Despite the advantages modelling techniques offer, they cannot be used for very large networks. The graphs are complex and hard to read.

This paper investigates security models, which consider risk assessment approaches to be applied for threat modelling, network hardening and risk analysis. Furthermore, we discuss the challenges related with the cost effective selection of countermeasures as well as clearly define research gaps in the area of risk assessment.

The remainder of the paper is organised as follows. In Section II, we present visualisation techniques, which often are applied for quantitative and qualitative risk assessment approach. In Section III we discuss security models, which apply risk assessment based methodologies to estimate risk. Section IV provides main challenges and limitations. In Section V we illustrate our research model

which is used to address limitations in the field. Finally we conclude in Section VI.

## II. INCREASING NETWORK SECURITY BY APPLYING VISUALISATION TECHNIQUES

### A. Visualisation Techniques

In order to identify and assess informational assets for threats, vulnerabilities and risks, and implement risk management practices to counteract them, it is necessary to have a clear picture of a state the network holds. Further to this, every security specialist can analyse the prospective adversaries and propose related security measures. With the complete picture, the risk and cost effective strategies can be used to make decisions, based not only on the expertise of the security specialist, but also based on data obtained through particular scenarios and examples [29].

Visualisation techniques can help to convert abstract data so that it can be more informative and easier to understand. The semantic analysis of information gathered from abstract data should provide knowledge that will support decision-makers in finding a solution on how systems should be protected. Being more practical, a graph which interconnects various aspects in a compact way is worth more than a long explanation.

In general, visualisation techniques are widely applied to analyse network security level and predict attacks, risks and possible threats. These techniques are attack graphs and attack trees. Attack graphs can represent all potential vulnerabilities and potential attack paths an attacker can take in order to reach the target. Furthermore attack graphs act as a tool in finding critical paths in large networks based on the threats and vulnerabilities identified. However, initial attack graph for large size networks was complex and visually not clear due to number of attack paths. Later, by enhancing the visualisation of the graph and proposing monotonicity concept, attack graphs became more scalable [30]. As a result, the layout of an attack graph can be adjusted to represent the real enterprise network.

Attack graphs are often applied for network hardening in order to effectively represent a prior knowledge about vulnerabilities, their dependencies and network connectivity. In a graph, each path is an exploit that can have undesirable impact on system (Figure 1a).

Attack trees are often applied for quantitative risk assessment analysis, because of the simplified way of visualising network nodes, attacks and their dependencies. According to [29] the concept of a tree can be used to analyse and assess the attributes of a security system, e.g., a probability of an attack success, assess the risks and quantify the costs of possible damage and defence controls. In terms of quantitative risk assessment, a tree structure can help to quantify lowest cost security countermeasure and the total cost of an attack.

For more information about visualisation techniques applied for network hardening, please refer to [10, 25, 29].

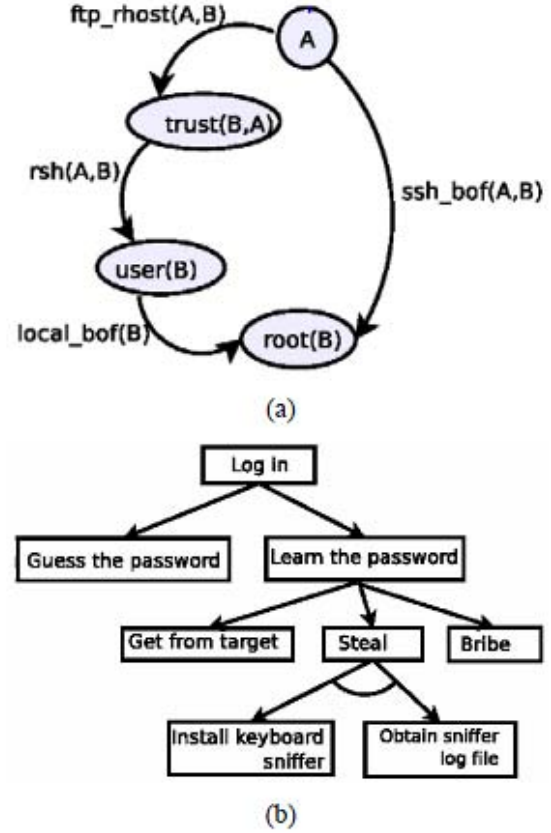


Figure 1. Attack scenarios by applying attack graph (a), attack tree (b)[29].

## III. SECURITY MODELS

Security of information assets is a priority for most organisations today. Investments are made to harden the perimeter, network devices, software bugs and to increase user awareness, however, to distinguish whether the investment was cost effective is a challenging task.

Researchers have developed a number of security models for valuating security investments and optimally investing into information security [11,12]. Differently from these models, the authors in [13, 21] have proposed a model which analyses optimal defence strategy as a finite game between an attacker and defender

In overall, security models can be classified based on the methodologies used to optimally invest into computer security. We have specified the following:

- Risk assessment models
- Cost-benefit models
- Game models
- Multi-objective decision support models

All of these methodologies help to distinguish how much should be invested into computer security, what benefits

an organisation would get from certain investment and optimal allocated resources, so an attacker would not have adverse effect on informational assets.

#### A. Risk Assessment Models

Risk assessment is one of the risk management procedures, which mainly provide guidelines in order to identify vulnerabilities, threats and their dependencies. The final goal of the process is to let decision makers be aware of possible risks that should be reduced, transferred or taken.

In general, risk assessment can be undertaken in two ways, by applying qualitative or/and quantitative analysis methods. Qualitative methods require a good knowledge in the areas of vulnerability assessment, threat analysis to correctly predict probabilities of various attacks and possible impacts on the assets [22]. Quantitative approaches offer mathematical methods for calculating risks, impacts and other relative cost factors.

We have made a comparison of risk assessment models analysed in the literature to help evaluate IT security investments. Table I summarises the main characteristics of the methodologies which can be used to calculate Return on Investment (ROI), Return on Attack (ROA) or to estimate the risk value, based on the input data, i.e. value of assets, threat likelihood, vulnerability severity.

The risk assessment models that result in lists of risks, ROI value or set of scenarios do not provide enough information to prioritize risks when multiple resource constraints are considered, e.g. budget or time constraints.

Another drawback of the simplistic risk assessment models relates to their non applicability to the realistic case scenarios. Assessing risks of information security is a very difficult step, because the data on the likelihood and costs associated with the risk factors is often limited and constantly changing. Thus, multi-objective conditions should be considered within the quantitative risk assessment process to optimally invest into network security.

#### B. Game Models

Differently from the models covered above, authors in [13][21] have proposed a game-theoretic method for optimal network hardening, where the overall cost for a defender is reduced based on the algorithm proposed. Game theoretical analysis is useful for decision, modelling and control processes related to network security.

An interaction between an attacker and a defender treated as a two-player stochastic game in which action sets, costs/reward functions and transition probabilities can be defined. By creating specific scenarios, a game theory provides information such as attacker's goal, steps he or she will follow, what are the rewards for each of the players, what the expenses are or how the resources are utilised during an attack [25]. Furthermore, game-models can help to estimate optimal network hardening strategies used to defend the attack in advance.

The main drawback of such theory lies in hypothetical assumptions, which may not be applicable in a real state of affairs. The steps an attacker may take are only predictable, so the final resource allocation methodology for a defender may be misleading.

TABLE I. COMPARISON OF RISK ASSESSMENT MODELS

References	Risk Assessment Models		
	Method	Input	Output
[23]	Quantitative	Threats, Success/Failure data, Asset value	Rik-vulnerability mapping, Risk index
[15]	Quantitative	Assets, Threats, Probabilities	ALE, Risk value, Total Damage
[2]	Hidden Markov method (HMM)	IDS alerts	Threat index Risk index
[7]	Quantitative	Asset values, Actual Threats, Potential Threats, Vulnerabilities, Frequency of threats	Total Risk, Risk for host index
[8]	Quantitative	Asset value, Relative ranking of vulnerabilities and threats Exposure factor (EF) ALE, ARO, SLE, Safeguard cost	Return on Investment (ROI)
[6]	Qualitative and Quantitative	An attack tree, SLE, ARO, ALE, EF, Safeguard cost, Cost of loss	ROI, Return on Attack(ROA)
[24]	Quantitative and Qualitative	Asset value, Severity of Vulnerability, Likelihood of an Attack, SLE, ALE, ARO Attack tree	ROI

#### C. Economic Models with the Cost-benefit Analysis

Cost-benefit analysis looks into intangible costs/returns and addresses time perspective. The simplicity of the frameworks can give suitable investment solutions for low risk investments. However, these methods do not consider uncertainty and give misleading indications for long term investments.

Arora et.al has proposed a risk management framework, which determines costs and benefits of information security solutions [14]. They evaluate investments based on the benefits each invested dollar of investment brings, as well as reduce expected loss or risk. Furthermore, a cost-benefit trade-off is identified in relation to risk based return on investment (RROI) rather than ROI. To calculate RROI, an incident risk is calculated first (Eq.1). To distinguish between how cost



effective the security solution is and how it will affect the incident risk, authors have introduced a bypass rate, which is estimated based on the effectiveness of it to an incident type (Eq. 1).

$$Risk = \frac{ObservedDamage(Incident\ type)}{Net\ bypass\ rate(Incident\ type)} \quad (1)$$

The model is rather hypothetical than practical because of the challenges in estimating bypass rate of the security solution and obtaining true costs of lost productivity or other possible attack consequences.

The cost-benefit analysis methodology proposed by Wei et.al, can be used to calculate the cost of detecting an intrusion and responding to it based on qualitative and quantitative risk management approaches. The overall goal of the methodology is to determine the trade-off between costs required to respond to the IDS events and benefits the investment brings. To calculate the total cost for the event the equation is as follows:

$$Cost_{total} = Progress * DamageCost + ResponseCost + OperationCost \quad (2)$$

From their model, a cost benefit trade-off is calculated and if the response cost is higher than the damage cost, IDS will not log an event; however, in case the response cost is lower than the damage cost, a response to an attack will be initiated. The model is not designed for real time intrusion detection and cost calculation is not effective for this case, however, this model can be used as a background for future developments.

#### D. Multi-objective Decision Support Models

Best security practice dictates that security requirements be based on risk assessment [17]. However, simplistic risk assessment that as a result lists risks based on the set of pre-defined scenarios does not provide sufficient information. Furthermore, when additional requirements are requested, such as time or budget constraints, these models cannot be applied.

Multi-attribute analysis help decision makers evaluate alternatives when conflicting objectives must be considered and balanced. Once constructed, a multi-attribute analysis framework also provides the basis from which decision makers can evaluate alternative risk-mitigation strategies [17].

The approach to match vulnerabilities to security profiles and enabling organisations to choose a minimal cost security profile providing maximal vulnerability coverage was analysed in [26]. The idea behind their approach is that any given security technology addresses only specific vulnerabilities and could possibly create additional vulnerabilities, named residual vulnerabilities.

The authors have made an assumption that if a known vulnerability is covered by a particular security technology, the risk of that vulnerability being exploited is uniform, which most probably in a real case scenario would not always be the truth, as the risk level depends upon the severity of the vulnerability and impact it possesses to a system. The multi-objective problem introduced is to minimise the residual vulnerabilities and cost of implementing security measures. The problem was simplified to a set-covering problem, solved using genetic algorithm (GA) adapted to the weighted sum fitness function (Eq. (3)). As an input, authors have used generic set of security policies capable of covering one or more generic vulnerabilities, previously proposed in [27].

$$F = \alpha \sum_{i=1}^m a_i r_i + \beta \sum_{j=1}^n c_j s_j \quad (3)$$

Where  $\alpha + \beta = 1$  and  $\alpha, \beta \leq 1$  represent preferences of the organisation. First objective maximises the coverage of vulnerabilities, or in other words, minimises the weighted residual vulnerability. Second objective minimises the costs to the organisation.

Based on the Butler's initial multi-attribute assessment framework, a multi-objective optimal security hardening problem was formulated in [28]. The authors have modelled the system administrator's decision problem as three optimisation problems on the attack tree model. The transformation of the problems has led to more cost-benefit solutions required by the decision maker. The multi-objective problem then is to find a vector of security measures which minimises the total security control cost and residual damage. To solve problems authors have used SGA and NSGA-II algorithms. From the results obtained by the NSGA-II, only few robust solutions with good balance between residual damage were obtained. The gaps in the Pareto front are signalling that authors have not considered dependencies between the security measures. This limits the ability to provide real life solutions when multiple objectives should be considered.

#### IV. CHALLENGES

Based on the state-of-the art work related to risk assessment approaches, when a decision maker has to consider a set of security countermeasures to be applied to increase the security and reduce the risk of possible losses, the main challenge is to combine conflicting factors into an analysis.

From the discussed work, the factors such as cost, risk, threats, likelihood are widely used, however, the issue like how the risk management can help in defining the risk and ways to reduce it, has not been raised before. Currently, there is no particular model introduced in the literature, which would combine general risk

the researchers to develop risk based models, which could

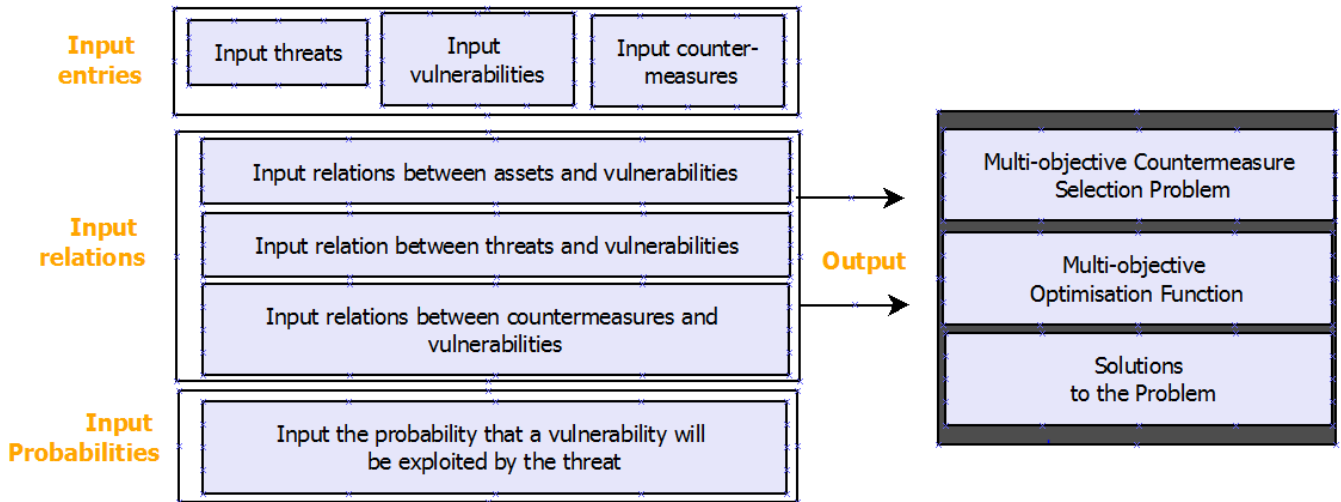


Figure 2. Multi-objective risk assessment model flowchart.

management factors into a multi-objective optimisation problem related to conflicting factors of cost and risk.

Furthermore, in order to be able to answer the question, like “What risk can I accept”, simplistic risk assessment models cannot be used. The following criteria should be considered:

- Multiple objectives, which would compound cost related factors.
- Be more practical.
- Consider the impact on confidentiality, integrity, availability (CIA).
- Use of tangible costs.

Considering the factors mentioned above, the model could help in making cost-effective decisions.

## V. RESEARCH RECOMMENDATIONS

Based on the challenges covered above, we have designed a model which in a coherent, structured way applies risk assessment procedure and optimisation routine for efficient search of cost effective solutions for multi-objective security countermeasure selection problem.

Figure 2 shows a flowchart of the research model we are working on, where input entries, relations and probabilities are used to formulate a multi-objective countermeasure selection problem and an optimisation function. Applying optimisation techniques, the process of selecting countermeasures can support with cost effective decisions.

## VI. CONCLUSION

The importance of appropriate security models and risk management procedures in modern society is well understood. There has been an increased interest between

help in dealing with the identification of threats, vulnerabilities and risks. Despite that, most of the models can only be used to compute effectiveness of investments in terms of calculating the ROI index or are limited by their applicability for real case scenarios.

Visualisation techniques open the ability to illustrate large data formats so that gained data would be used for knowledge and improved awareness between information system users. Furthermore, use of visualisation techniques helps in identification of threats in networked systems.

## REFERENCES

- [1] C.Maple, P. Phillips, “UK Security Breach Investigations Report”, 7Safe, United Kingdom, 2010.
- [2] Jie Ma; Zhi-tang Li; Hong-wu Zhang; , "A Fusion Model for Network Threat Identification and Risk Assessment," *International Conference on Artificial Intelligence and Computational Intelligence*, 2009. *AICI '09.*, vol.1, pp.314-318.
- [3] M.Templeman, M. Beishon, L. Malachowski, A.Wilson, T. Nash, L. Robertson, Information security - best practice measures for protecting your business, Tech. rep., Department of Trade and Industry (2005).
- [4] ISO/IEC 27001:2005, Information Technology - Security techniques – Information Security Management Systems - Requirements, International Organisation for Standardization (2005).
- [5] ISO/IEC 17799:2005, Information technology - Code of practice for information security management (2005).
- [6] S. Bistarelli, F. Fioravanti, P. Peretti, “Defense Trees for Economic Evaluations of Security Investment” , *ARES'06*, pp.416-423.



- [7] H. Lv, "Research on Network Risk Assessment Based on Attack Probability", *International Workshop on Computer Science and Engineering*, vol. 2, 2009, pp.376–381.
- [8] A. Asosheh, B. Dehmoubed, A. Khani, "A new Quantitative Approach for Information Security Risk Assessment", *International Conference on Computer Science and Information Technology* 2009, pp. 222–227.
- [9] L. Wang, S. Noel, S. Jajodia, "Minimum-cost network hardening using attack graphs", *Computer Communications* (2006), vol.29, pp.3812–3824.
- [10] C.Maple, V.Viduto,"A Visualisation Technique for the Identification of Security Threats in Networked Systems", *In proceedings of Information Visualisation*, 2010, pp.551-556.
- [11] T. Bandyopadhyay, "Information Security Investment in Prevention and Detection Regimes – Towards an Aggregate Economic Model," *Proceedings of SAIS 2007*.pp. 142 – 147.
- [12] A. L. Couch, N. Wu, H. Susanto, "Toward a Cost Model for System Administration", *Proceedings of the Niteenth Large Installation System Administration Conference (LISA 05)*,2005, pp. 125-141.
- [13] W.Jiang H. Zhang, Z. Tian, X. Song, "A Game Theoretic Method for Decision and Analysis of the Optimal Active Defense Strategy, *In Proceedings of the 2007 International Conference on Computational Intelligence and Security (CIS '07)*, 2007, pp. 819-823.
- [14] A.Arora, D. Hall, C. Pinto, D. Ramsey, R. Telang, "Measuring the Risk-based Value of IT Security Solutions", *IT Professional* vol.6, 2004, pp. 35–42.
- [15] A.Ekelhart, S.Fenz, M. Klemen, E. Weippl, "Security Ontologies: Improving Quantitative Risk Analysis, in "40th Annual Hawaii International Conference on System Sciences", 2007.
- [16] H.Wei, D. Frinke, O. Carter, C. Ritter, "Cost-benefit Analysis for Network Intrusion Detection systems", in 'CSI 28th Annual Computer. Security Conference', 2001.
- [17] S. A. Butler, P. Fischbeck, " Multi-attribute risk assessment", Tech. Report, *Proceedings of SREIS01*, 2001.
- [18] Gupta, M., Rees, J., Chaturvedi, A. & Chi, J. "Matching Information Security Vulnerabilities to Organizational Security Profiles: a Genetic Algorithm Approach", *Decision Support Sysemst*, vol.41(3), (2006), pp.592–603.
- [19] R. Dewri, N. Poolsappasit, I. Ray, D. Whitley, Optimal security hardening using multi-objective optimization on attack tree models of networks, in 'Proceedings of the 14th ACM conference on Computer and communications security', ACM, 2007 , pp. 204–213.
- [20] T.Neubauer, C.Stummer, E.Weippl, "Workshop-based Multi-objective Security Safeguard Selection", *International Conference on Availability, Reliability and Security*, 2006, pp. 366-373.
- [21] W. Jiang, B. Fang, H. Zhang, Z. Tian, X. Song, "Optimal Network Security Strengthening Using Attack-Defense Game Model," *Third International Conference on Information Technology: New Generations*, , 2009 , pp. 475-480.
- [22] G.Y.Liao, C.H.Song, "Design of a Computer-aided System for Risk Assessment on Information Systems", *IEEE 37<sup>th</sup> Annual International Carnahan Conference on Security Technology*, 2003, pp. 157-162.
- [23] S. Kondakci, "A Composite Network Security Assessment", *Proceedings of the Fourth International Conference on Information Assurance and Security*, 2008, pp. 249-254.
- [24] M. Ketel, "IT security risk management", *Proceedings of the 46th Annual Southeast Regional Conference*, 2008, pp. 373-376.
- [25] V. Viduto, C. Maple, W. Huang, "An Analytical Evaluation of Network Security Modelling Techniques Applied to Manage Threats," *International Conference on Broadband, Wireless Computing, Communication and Applications*,2010, pp. 117-123.
- [26] M. Gupta, J. Rees, A. Chaturvedi, J. Chi, "Matching Information Security Vulnerabilities to Organizational Security Profiles: a Genetic Algorithm Approach', *Decis. Support Syst.* vol.41(3), 2006, pp. 592–603.
- [27] R. H. Anderson, P. M. Feldman, S. Gerwehr, B. Houghton, R Mesic, J. D Pinder, J. Rothenberg, J. Chiesa, "Securing the U.S. Defense Information Infrastructure: A Proposed Approach", Technical report, RAND corporation, 1999.
- [28] R. Dewri, N. Poolsappasit, I. Ray, D. Whitley, "Optimal Security Hardening Using Multi-objective Optimization on Attack Tree Models of Networks", in 'CCS '07: Proceedings of the 14th ACM conference on Computer and communications security', 2007, pp. 204–213.
- [29] V. Viduto, C. Maple, W. Huang, "Managing Threats by the Use of Visualisation Techniques", *Int. J. Space-Based and Situated Computing*, Vol. 1, Nos. 2/3, pp. 204–212.
- [30] Jha, S., Sheyner, O. and Wing, J. "Two formal analysis of attack graphs", in *CSFW '02: Proceedings of the 15th IEEE Workshop on Computer Security Foundations*, 2002, p. 49.

# A Complete Planner Design of Microstrip Patch Antenna for a Passive UHF RFID Tag

Tashi<sup>1,2</sup>, Mohammad S. Hasan<sup>2</sup> and Hongnain Yu<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering  
College of Science and Technology, Royal University of Bhutan  
Rinchending, Phuentsholing, Bhutan

<sup>2</sup>Faculty of Computing, Engineering and Technology  
Staffordshire University

Becondside, Stafford, ST18 0AD, England, UK  
<sup>1</sup>tashi@cst.edu.bt, <sup>2</sup>{m.s.hasan, h.yu}@staffs.ac.uk

**Abstract**—A micro-strip patch antenna for a passive radio frequency identification (RFID) tag which can operate in the ultra high frequency (UHF) range from 865 MHz to 867 MHz is presented in this paper. The proposed antenna is designed and suitable for tagging the metallic boxes in the UK and Europe warehouse environment. The design is supplemented with the simulation results. In addition, the effect of the antenna substrate thickness and the ground plane on the performance of the proposed antenna is also investigated. The study shows that there is little affect by the antenna substrate thickness on the performance.

**Keywords**—RFID; Passive Tag ; Tag antenna; Micro-strip patch antenna

## I. INTRODUCTION

Radio frequency identification (RFID) technology [1] is growing tremendous demand in the supply chain management system. RFID is an automatic identification (Auto ID) technology [2]. It is a pervasive computing technology for collecting and gathering data from a tagged item. The data is stored in the mobile device called tag. When the tag comes in the reader's reading zone, the data is collected by the reader without any need of physical contact. The data in the tag may be the identification number, location information, or specification of product such as price, brand, date, etc. Unlike bar code technology, the RFID technology does not require light-of-sight and reads longer distance [3]. Such advantages help the supply chain to operate very fast and efficiently.

Recently, a passive RFID system that operates in ultra high frequency (UHF) band from 860MHz to 960 MHz [4] is getting considerable attention. Because passive tags are very cheap due to absence of onboard battery and using UHF band can provide longer reading range, high data rate and small sized antenna.

Currently, a label-type dipole antenna is commonly used as a tag antenna for a passive UHF RFID tag and it is printed on a very thin film at low cost [5, 6] to reduce the overall cost of the tag. The commercial passive UHF RFID tags are shown in Fig. 1. The passive tags are embedded in cardboard boxes, ID card, airline baggage strip, passport, clothing tags, etc. However, papers [7] and [8] have reported that these tags undergo a serious performance degradation when it is mounted or attached

onto the metallic surface or in the presence of water. This is because their performances are optimised in the free space. When passive tags are mounted on the metallic surface, label-type dipole tag antenna is short-circuited by the metallic surface. This results in changes in the performance parameters such as radiation pattern, antenna impedance, gain and bandwidth of the RFID tag antenna. Therefore, when a passive tag is placed onto the metallic surface, it fails to be read by the reader within the normal reading zone. On other hand, there are still increasing demands on RFID technology in the supply chain application for tagging the metallic objects since most of the items in the wire house are made of metal or encased the metallic boxes or container.

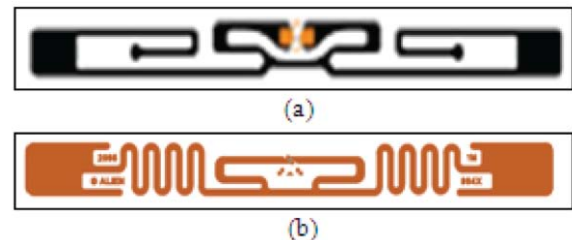


Figure 1. Commercial Passive UHF RFID tags (a) Avery Dennison tag [9] and (b) Alien Squiggle tag [10].

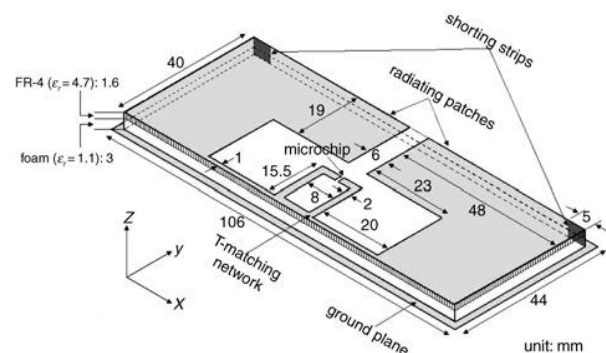


Figure 2. A micro-strip patch antenna that the antenna trace is shorted to the ground plane [11].

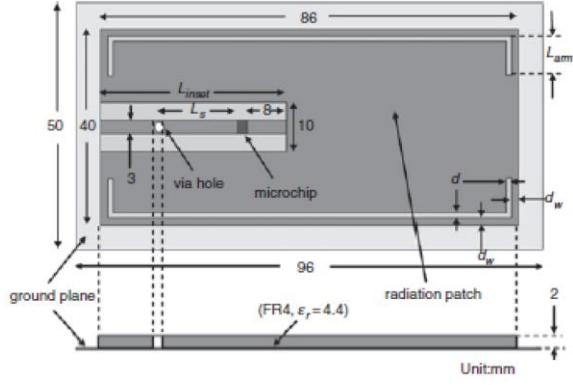


Figure 3. A micro-strip patch antenna that the feeding is shorted to the ground plane [12].

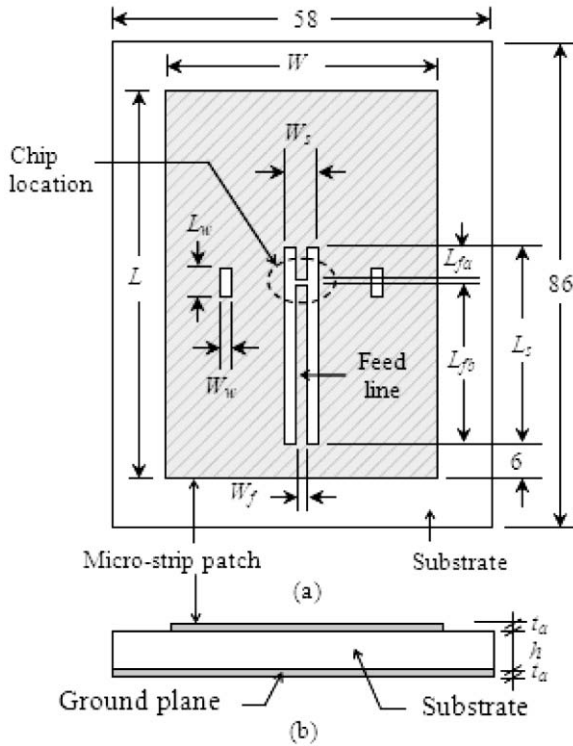


Figure 4. Geometry of the proposed micro-strip patch antenna (a) top view and (b) side view.

This problem may be solved by using an active RFID tag, but it will be very costly. The alternative solutions are by keeping enough separation gaps between the tag and the metallic surface or by designing a tag antenna that can operate using the ground plane. Micro-strip patch antenna and planar inverted-F antenna (PIFA) are attractive choices as both antennas use the ground plane. From the literature review, RFID tags are employed using the micro-strip patch antenna in [11-18] and the PIFA in [19-21]. All these tag antennas have an electrical connection between the antenna trace and the ground plane (e.g., [11] shown in Fig. 2). In some cases there is an electrical connection between the feeding line and the ground plane (e.g., [12] shown in Fig. 3). This leads to the cross-layered construct of tag antennas. In some PIFA, the tag antenna

has multilayered construction (e.g., [20]). Such design presents complex and costly antenna structure. For small size, easy installation, lower cost and easy manufacturing in bulk, tag antennas are preferred to be completely planar so that it can be directly printed on the antenna substrate without having any cross and multilayered construction.

In this paper, a micro-strip patch antenna is proposed as a tag antenna for a passive UHF RFID tag that can be used for tagging the metallic containers or metallic boxes in the warehouse environment. The fundamental characteristic of the proposed antenna is that the antenna trace (patch) is suspended above the ground plane without any electrical connection between the antenna trace and the ground plane. This eliminates the cross and multilayered construction without compromising its performance. Therefore, the proposed antenna is completely planar with low antenna profile and easy to fabricate. The rest of the paper is organised as follow. Section II discusses the proposed tag antenna design. Section III presents the simulation results of the proposed antenna. Finally, section IV addresses the conclusion.

## II. THE PROPOSED TAG ANTENNA DESIGN

The proposed micro-strip patch antenna for a passive UHF RFID tag is considered to operate in the UK and Europe warehouse environment for tagging the metallic boxes or metallic containers. The tag shall be easily tunable from 865 MHz to 867 MHz range which is UHF RFID frequency band for the UK and Europe. The operating frequency of proposed antenna is selected as 866 MHz which is the centre frequency of the band.

Passive tags consist of a chip and an antenna. In this proposed design, the Alien Higgs-3 [22] for EPC Class 1 Gen 2 RFID tag chip is selected. The Alien Higgs-3 exhibits an impedance of  $Z_c = (31 - j212) \Omega$  at 866 MHz resonance frequency and requires a minimum of -14 dBm power to turn on the chip. In order to deliver the maximum power from the antenna to the chip, the input impedance of the antenna,  $Z_a$  should be complex conjugately matched to the chip impedance,  $Z_c$  (i.e.  $Z_c = Z_a^*$ ). Therefore, the proposed antenna is designed for  $Z_a = (31 + j212) \Omega$ .

The geometry of the proposed antenna is shown in Fig. 4. The top layer is antenna trace and bottom layer is the antenna ground plane. Both layers are copper with the same thickness ( $t_a$ ) of 0.0358 mm. The middle layer is FR4 substrate with thickness ( $h$ ) of 1.6 mm, dielectric constant ( $\epsilon_r$ ) of 4.9 and loss tangent ( $\delta$ ) of 0.025. The FR4 substrate and ground plane has the same length and width of 86 mm and 58 mm, respectively. The antenna trace (patch) is slotted in rectangular shape with slot length  $L_s = 40$  mm and slot width  $W_s = 6$  mm. The rectangular slot is particularly for inserting the feeding line inside the patch to reduce the size of tag antenna. There are also two small slots on the patch which are called window and these have length  $L_w = 6$  mm and width  $W_w = 2$  mm. The window is designed for fine tuning of the patch resonance frequency at the desired operating frequency without changing other parameters of the patch.

Traditional micro-strip patch antenna is single feed that can be designed as direct feed type or coupled feed type and keeping reference with respect to the ground plane [23]. Such feeding methods require a cross-layered construction in order to attach the chip. To avoid cross-layered structure, T-match is proposed in this design. The T-match feeding lines are inserted inside the rectangular patch slot. There are no electrical connection between the patch and the ground plane. The chip location is fixed at the centre of the patch (i.e. at the origin of principle x-y plane of the proposed antenna). The RF port-1 of the RFID chip is connected to the patch with feeding line length  $L_{fa} = 6.5$  mm and the RF port-2 (i.e. ground) is connected to the patch with feeding line length  $L_{fb} = 32.5$  mm. Both feeding line has width  $W_f = 2$  mm. The input impedance of the proposed antenna is tuned by changing geometry parameters of T-match feeding line.

### III. SIMULATION RESULTS OF THE PROPOSED TAG ANTENNA

The proposed micro-strip patch antenna design is simulated in Sonnet Lite version 12.52, electromagnetic (EM) simulator [24] and in AWR Design Environment version 9.04 [25] and the results are compared for validation. Both EM simulators work based on the method of moments (MoM).

#### A. Input Impedance

Fig. 5 and Fig. 6 show the simulated input resistance and input reactance, respectively against UHF frequency of the proposed micro-strip patch antenna. Both EM simulators results are closely matched.

#### B. Return Loss

Fig. 7 shows the return loss of the proposed micro-strip patch antenna at 866 MHz resonance frequency. The Sonnet simulation results also compared to the AWR simulation result and both show strong agreement. The minimum value of the simulated return loss ( $S_{11}$ ) at the resonance frequency from the Sonnet is -21.03 dB. The half-power bandwidth (return loss < -3 dB) is 46 MHz (5.31%), from 843.50 MHz to 889.50 MHz. The simulated < -10 dB bandwidth of the proposed tag antenna is 15 MHz (1.73%), from 858.50 MHz to 873.50 MHz. Both bandwidths satisfy the design goal of the proposed tag antenna which is 2 MHz (i.e. from 865 MHz to 867 MHz) as well as meet the requirement bandwidth (500 kHz) of the ISO/IEC 18000-7 standard. Fig. 8 shows the current distribution over the proposed patch antenna at 866 MHz operating frequency.

#### C. Radiation Patterns

Fig. 9 shows the simulated radiation patterns of the proposed micro-strip patch antenna at 866 MHz in the E-plane (x-z plane). It is also called as elevation plot in 2D. The 3D radiation pattern of the micro-strip patch antenna will look like dome shape. Fig. 10 shows the simulated radiation patterns of the proposed micro-strip patch antenna at 866 MHz in the H-plane (x-y plane). It is also called as azimuth plot. At zero degree, E-plane has a maximum radiation of 3.28 dB and H-plane has maximum

radiation of 3.29 dB. The radiation patterns are almost omnidirectional in the E-plane and almost bidirectional in the H-plane.

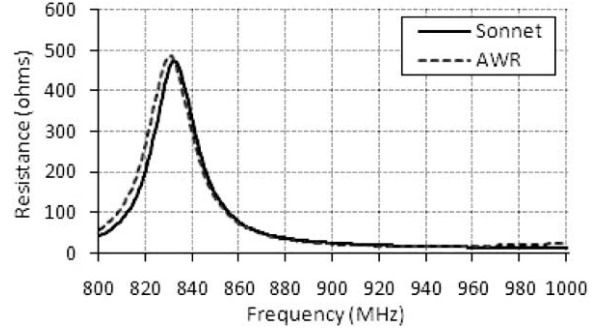


Figure 5. Simulated input resistance against UHF frequency for the proposed micro-strip patch antenna.

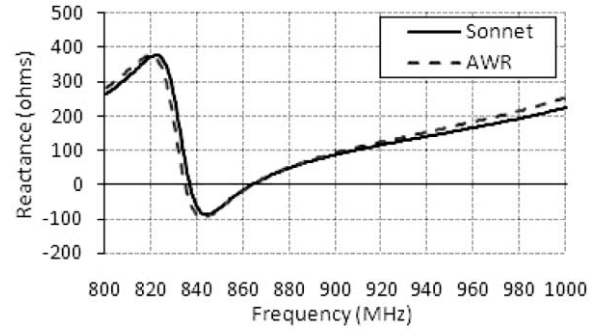


Figure 6. Simulated input reactance against UHF frequency for the proposed micro-strip patch antenna.

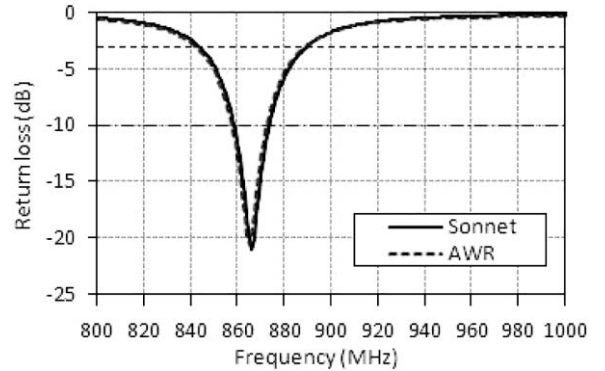


Figure 7. Simulated return loss against UHF frequency for the proposed micro-strip patch antenna.

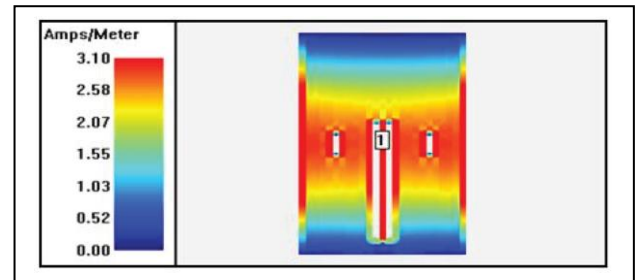


Figure 8. Simulated current distribution over the proposed micro-strip patch antenna at 866 MHz resonance frequency.



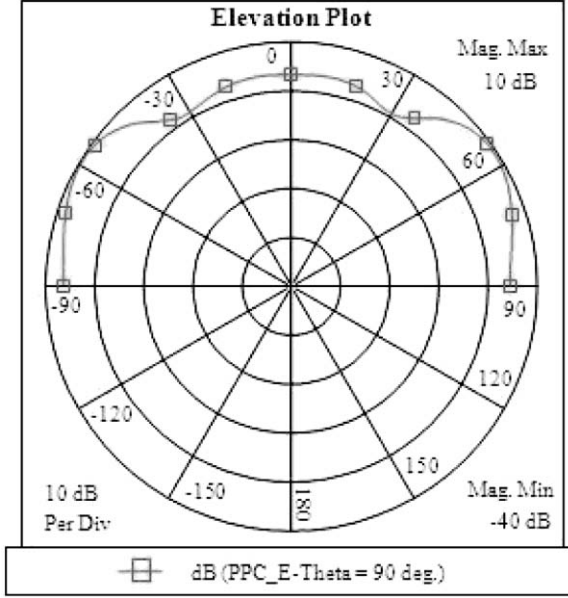


Figure 9. Simulated radiation patterns at 866 MHz for the proposed micro-strip patch antenna in E-plane.

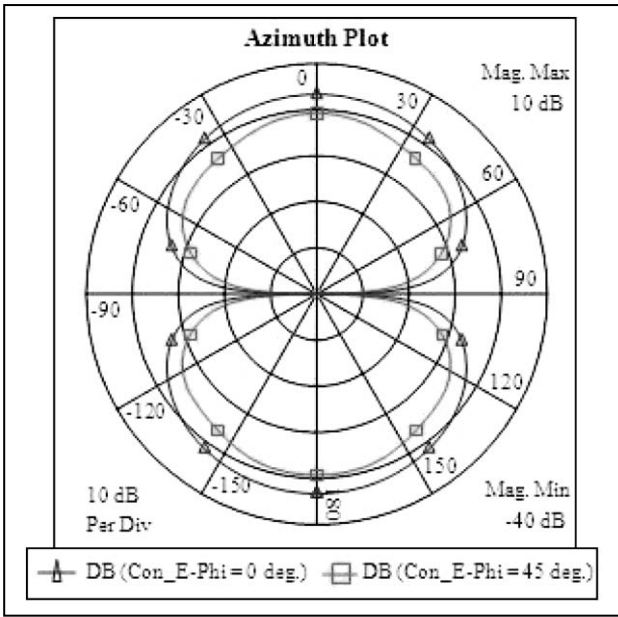


Figure 10. Simulated radiation patterns at 866 MHz for the proposed micro-strip patch antenna in H-plane.

#### D. Effect of Substrate Thickness

Fig. 11 shows the simulated return loss against UHF frequency of the proposed micro-strip patch antenna for different thickness ( $h$ ) of substrate while keeping other parameters the same. The proposed antenna exhibits an optimised performance for  $h = 1.6$  mm. It shows that when the antenna substrate thickness ( $h$ ) increases, the return loss is significantly increased and thereby reduces the performance of the antenna. So the selection of the substrate thickness for micro-strip patch antenna is very important.

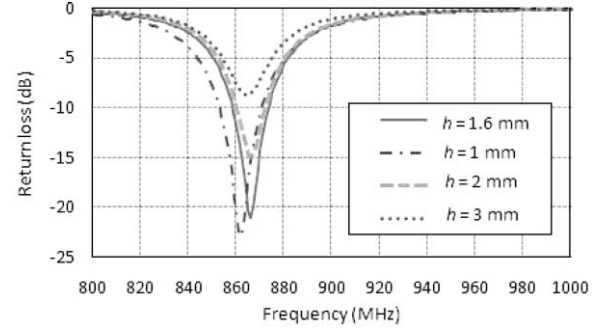


Figure 11. Simulated return loss of the proposed micro-strip patch antenna for different thickness ( $h$ ) of the substrate while keeping other parameter same.

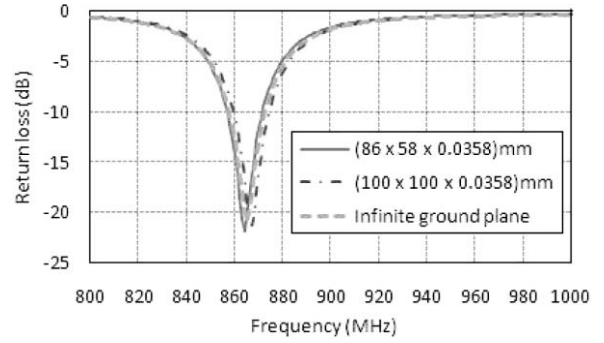


Figure 12. Simulated results return loss of the proposed micro-strip patch antenna for different size of the ground plane while keeping other parameters same.

#### E. Effect of Ground Plane

The antenna is designed on the infinite ground plane and substrate in both the Sonnet and AWR EM simulator environment. When the infinite ground plane and substrate are replaced by the finite ones in real world application, there are affects of the ground plane on the radiation patterns and resonance frequency of the micro-strip patch antenna [15, 19]. The effect of the ground plane on the return loss of the proposed tag antenna is studied by using three different ground plane sizes e.g. 86 mm x 58 mm x 0.0358 mm, 100 mm x 100 mm x 0.0358 mm and infinite ground plane. The substrate length and breadth are taken as same as ground plane with the thickness of 1.6 mm in all the cases. Fig. 12 shows the simulated return loss against UHF frequency of the proposed micro-strip patch antenna while keeping other parameters the same. The change in bandwidth and shift in resonance frequency due to different ground planes are presented in Table I. The change in bandwidth is evaluated based on the half power ( $< -3$  dB) and  $< -10$  dB return loss bandwidth. There is not much effect on the bandwidths but there are shifts in the resonance frequencies. However, the entire range of shifts in the resonance frequencies is within the desired goal of the proposed tag antenna.

TABLE I. SIMULATED RETURN LOSS AND BANDWIDTH FOR DIFFERENT SIZE OF THE GROUND PLANE.

Return loss (dB)	Ground plane size (mm)	Frequency range (MHz)	Bandwidth (MHz)	Resonance frequency (MHz)	Minimum return loss (dB)
- 3	86 x 58 x 0.0358	843.5 – 889.5	46	866	-21.03
	100 x 100 x 0.0358	843.82 – 890.51	46.69	867.12	-21.06
	Infinite	841.34 – 887.28	45.94	864	-21.86
- 10	86 x 58 x 0.0358	858.5 – 873.5	15	866	-21.03
	100 x 100 x 0.0358	859.44 – 874.51	15.04	867.12	-21.06
	Infinite	856.71 – 871.36	14.65	864	-21.86

#### IV. CONCLUSION AND FUTURE WORK

Micro-strip patch antenna is proposed as tag antenna for a passive UHF RFID tag. The proposed tag is designed for tagging the metallic containers in the UK and European warehouse environment. The proposed antenna is designed to operate at centre frequency of 866 MHz and it is tunable from 865 MHz to 876 MHz UHF frequency band. The design is supplement by the simulation results. The paper also presents the effect of the substrate height and the ground plane on the antenna performance. The maximum return loss of the proposed antenna loss (S11) is -21.03 dB. The half-power bandwidth (return loss < -3 dB) is 46 MHz (5.31%), from 843.50 MHz to 889.50 MHz. The simulated < -10 dB bandwidth of the proposed antenna is 15 MHz (1.73%), from 858.50 MHz to 873.50 MHz. Both return losses bandwidth satisfy the design goal. In the future work, the micro-strip antenna for a passive UHF RFID shall be fabricated using the design proposed in this paper. Then the prototype micro-strip patch antenna will be implemented and analyse its read range when attached to the metallic surface.

#### ACKNOWLEDGMENT

This work was supported by the eLINK (east-west Link for Innovation, Networking and Knowledge Exchange) project.

#### REFERENCES

- [1] R. Want, "An introduction to RFID technology," *IEEE Pervasive Computing*, vol. 5, pp. 25-33, 2006.
- [2] R. Want, *RFID Explained: A Primer on Radio Frequency Identification Technologies*, First ed.: Morgan & Claypool, 2006.
- [3] K. Finkenzeller, *RFID Handbook Fundamentals and Applications in Contactless Smart Cards and Identification*. Chichester: John Wiley and Sons Ltd, 2003.
- [4] G. Backhouse. (2006). RFID: Frequency, standards, adoption and innovation. *JISC Technology and Standards Watch*. [Online]. Available: <http://www.jisc.ac.uk/media/documents/techwatch/tsw0602.pdf> [Accessed: Jan. 2010]
- [5] K. V. S. Rao, *et al.*, "Antenna design for UHF RFID tags: a review and a practical application," *IEEE Transactions on Antennas and Propagation*, vol. 53, pp. 3870-3876, 2005.
- [6] C. Cho, *et al.*, "Broadband RFID tag antenna with quasi-isotropic radiation pattern," *Electronics Letters*, vol. 41, pp. 1091-1092, 2005.
- [7] D. M. Dobkin and S. M. Weigand, "Environmental effects on RFID tag antennas," in *Microwave Symposium Digest, 2005 IEEE MTT-S International*, 2005, p. 4 pp.
- [8] J. D. Griffin, *et al.*, "RF Tag Antenna Performance on Various Materials Using Radio Link Budgets," *Antennas and Wireless Propagation Letters, IEEE*, vol. 5, pp. 247-250, 2006.
- [9] AveryDennison. *RFID Products*. [Online]. Available: <http://www.rfid.averydennison.com/products.php#2> [Accessed: Jan. 2010]
- [10] Alien. (2010) *ALN-9640 Squiggle Inlay*. [Online]. Available: [http://www.alientechnology.com/docs/products/DS\\_ALN\\_9640.pdf](http://www.alientechnology.com/docs/products/DS_ALN_9640.pdf) [Accessed: June 2010]
- [11] L. Xu, *et al.*, "UHF RFID tag antenna with broadband characteristic," *Electronics Letters*, vol. 44, pp. 79-80, 2008.
- [12] L. Mo, *et al.*, "Broadband UHF RFID tag antenna with a pair of U slots mountable on metallic objects," *Electronics Letters*, vol. 44, pp. 1173-1174, 2008.
- [13] K. H. Kim, *et al.*, "Fork-shaped RFID tag antenna mountable on metallic surfaces," *Electronics Letters*, vol. 43, pp. 1400-1402, 2007.
- [14] H. W. Son, *et al.*, "Design of wideband RFID tag antenna for metallic surfaces," *Electronics Letters*, vol. 42, pp. 263-265, 2006.
- [15] L. Ukkonen, *et al.*, "Effects of metallic plate size on the performance of microstrip patch-type tag antennas for passive RFID," *IEEE Antennas and Wireless Propagation Letters*, vol. 4, pp. 410-413, 2005.
- [16] C. Horng-Dean and T. Yu-Hung, "Low-Profile Meandered Patch Antennas for RFID Tags Mountable on Metallic Objects," *Antennas and Wireless Propagation Letters, IEEE*, vol. 9, pp. 118-121, 2010.
- [17] C. Horng-Dean and T. Yu-Hung, "Broadband Capacitively Coupled Patch Antenna for RFID Tag Mountable on Metallic Objects," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 489-492, 2010.

- [18] K. V. S. Rao, *et al.*, "UHF RFID tag for metal containers," in *Microwave Conference Proceedings (APMC), 2010 Asia-Pacific*, 2010, pp. 179-182.
- [19] M. Hirvonen, *et al.*, "Planar inverted-F antenna for radio frequency identification," *Electronics Letters*, vol. 40, pp. 848-850, 2004.
- [20] H. Kwon and B. Lee, "Compact slotted planar inverted-F RFID tag mountable on metallic objects," *Electronics Letters*, vol. 41, pp. 1308-1310, 2005.
- [21] C. Horng-Dean and T. Yu-Hung, "Low-Profile PIFA Array Antennas for UHF Band RFID Tags Mountable on Metallic Objects," *IEEE Transactions on Antennas and Propagation*, vol. 58, pp. 1087-1092, 2010.
- [22] Alien. (2010) *RFID ICs*. [Online]. Available: [http://www.alientechnology.com/docs/products/DS\\_H3.pdf](http://www.alientechnology.com/docs/products/DS_H3.pdf) [Accessed: Dec. 2010]
- [23] C. A. Balanis, *Antenna theory : analysis and design*, 3rd ed. ed. Hoboken, N.J.: [Great Britain] : Wiley-Interscience, 2005.
- [24] C. Blair and J. C. Rautio, "RFID design using EM analysis," in *Applications and Technology Conference (LISAT), 2010 Long Island Systems*, 2010, pp. 1-6.
- [25] AWR. (na) *Microstrip Patch Antenna*. [Online]. Available: [https://awrcorp.com/download/faq/english/examples/Microstrip\\_Patch\\_Antenna.aspx](https://awrcorp.com/download/faq/english/examples/Microstrip_Patch_Antenna.aspx) [Accessed: January 2011]



# A fast and effective way to improve the merging accuracy of multi-view point cloud data

Feng Li, Andrew Longstaff, Simon Fletcher, Alan Myers  
Centre for Precision Technologies, School of Computing & Engineering  
Huddersfield University  
Queensgate, Huddersfield, HD1 3DH, UK  
Feng.Li@hud.ac.uk

**Abstract**—In reverse engineering, in order to meet the requirements for model reconstruction, it is often necessary to locate and merge the different-view-measured cloud data in a global coordinate system. Many merging methods have been proposed, the method of three datum points is one of them and the registration precision of model data depends on the precision of three datum points which are selected. This paper introduces a new development of the “centroid of apexes” method instead of the former datum points to improve the three points positioning algorithm, the effectiveness of the methods is validated with experimental results and a revised algorithm is presented.

**Keywords**—reverse engineering; points registration; coordinate transform; reference marker; 3-D pointsets registration

## NOMENCLATURE

$p \ q$	Coordinate of feature points
$V \ W$	Vector between points
$v \ w$	Unit vector
$\begin{bmatrix} v \\ w \end{bmatrix}$	Unit vector matrix
$P$	Coordinate of any point
$R$	Rotation matrix
$T$	Translation vector
$\varepsilon$	Absolute error of two edges
$\Delta$	Relative error

## I. INTRODUCTION

The applications of three-dimensional (3D) shape measurement are widely used in the fields of industrial design and manufacturing, relic restoration, biomedicine and computer vision. There are various non-contact optical instruments involved in 3D surface measurement which are based on time-of-flight lasers [1], laser scanning [2], stereovision [3], and structured light [4]. These optical instruments can efficiently capture dense point clouds, which reveal the detail surface shape of the object being scanned. However, all of them can only obtain partial area of the object at one standpoint due to obstructions and the limited field of view of the sensor. In order to build a complete 3D model, it needs to collect point clouds acquired from different views. These multi-view scans are represented in their own local coordinate system, and geometrically aligning them to a global coordinate system is called the “registration problem”.

Solutions that are commonly used in practice for registration of multi-view Point Cloud include using datum markers, exploiting mechanical devices like turntables [5] or multi-joint robotic arms [6]. The markers can be planar or solid and are usually adhered on or near the object to be scanned. While the measuring sensor is taking point clouds from a specific view, the 3D coordinates of the markers within the view are obtained at the same time. The relative position and orientation of two scans can be easily determined if only three or more pairs of markers are visible in both views. This registration method is usually fast and reliable. However, except for the preparation work before the measurement, the drawbacks of this strategy include that the areas covered by the markers cannot be digitized reliably. This problem is outstanding especially for objects with small size and abundant details. Moreover, adhering markers on the surface is obtrusive or even prohibited in some applications.

## II. REGISTRATION ALGORITHM BASED ON 3-D POINT SETS METHOD

Coordinate transformation of 3D graphics includes geometric transformations of translation, proportion, rotation and shear. The data alignments in this paper are only translation and rotation transformation. Since three points can express a complete coordinate, multi-view data transformation will be achieved simply with three different reference points. Besl and McKay described manifold 3-D shape registration methods, including 3-D point sets, free-form curves and surfaces [7]. Among these methods, the 3-D point sets registration method is used in most places, especially in reverse engineering where the object shape is described as 3-D scan point sets. To carry out 3-D point sets registration, first construct least-square distance object function between the corresponding points, solve the object function based on the quaternions and the singular value decomposition (SVD) the rotation and the translation of the rigid movement [8-10].

Measurement data registration can be seen as a kind of rigid body movement, so a three-point alignment coordinate transformation method can be used to deal with data registration. Because three points can establish a coordinate, we can set up three datum points of the sphere center at a different view for the data alignment. The data registration of 3D measurement data will be achieved through the alignment of three datum sphere center points. In fact, the data alignment problem is converted to coordinate transformation.

The method of three-point alignment coordinate transformation: suppose datum feature points are  $p_1, p_2$  and  $p_3$ . The coordinates of the three datum points in the second measurement turn into  $q_1, q_2$  and  $q_3$ . Coordinate transformation can be achieved via three steps as derived by Mortenson and presented here for clarity [11]:

1. Transform  $p_1$  to  $q_1$ ;
2. Transform vector  $(p_2 - p_1)$  to  $(q_2 - q_1)$  (taking only the direction into consideration);
3. Transform the plane containing the three points  $p_1, p_2$  and  $p_3$  to the plane containing the three points  $q_1, q_2$  and  $q_3$ . The algorithm is:

Step 1: Set up vector  $(p_2 - p_1), (p_3 - p_1), (q_2 - q_1)$  and  $(q_3 - q_1)$  (1)

Step 2: Define  $V_1 = p_2 - p_1, W_1 = q_2 - q_1$  (2)

Step 3: Set up vector  $V_3$  and  $W_3$

$$\begin{cases} V_3 = V_1 \times (p_3 - p_1) \\ W_3 = W_1 \times (q_3 - q_1) \end{cases} \quad (3)$$

Step 4: Set up vector  $V_2$  and  $W_2$

$$\begin{cases} V_2 = V_3 \times V_1 \\ W_2 = W_3 \times W_1 \end{cases} \quad (4)$$

Obviously, the vector  $V_1, V_2$  and  $V_3$  constitute right-handed orthogonal lines, and the vector  $W_1, W_2$  and  $W_3$  also constitute right-handed orthogonal lines.

Step 5: Set up unit vector

$$\begin{cases} v_1 = \frac{V_1}{|V_1|}, v_2 = \frac{V_2}{|V_2|}, v_3 = \frac{V_3}{|V_3|} \\ w_1 = \frac{W_1}{|W_1|}, w_2 = \frac{W_2}{|W_2|}, w_3 = \frac{W_3}{|W_3|} \end{cases} \quad (5)$$

Step 6: Transform any point  $P_i$  in the system  $[v]$  to the system  $[w]$ , with transformation formula:

$$P_i^* = P_i R + T; \quad (6)$$

Step 7: As  $[v]$  and  $[w]$  are unit vector matrixes,  $[w] = [v]R$ , so the unknown rotation matrix about the  $[w]$  system is

$$R = [v]^{-1}[w]; \quad (7)$$

Step 8: Define  $P_1' = q_1$  and  $P_1 = p_1$ , put them into the equation, then the translation vector  $T$  is obtained;

$$T = q_1 - p_1[v]^{-1}[w]; \quad (8)$$

Step 9: Equation is rewritten:

$$P^* = P[v]^{-1}[w] - p_1[v]^{-1}[w] + q_1; \quad (9)$$

Fig.1 shows a three-point coordinates transformation

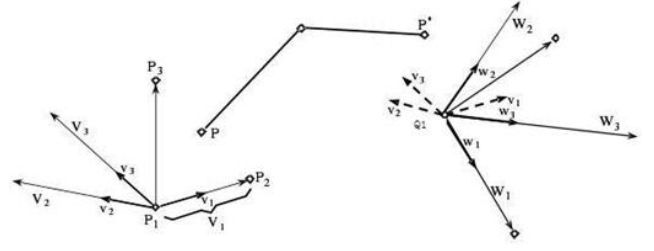


Figure 1. Three points to three points transformation.

Using the above re-positioning algorithm and introducing the reference points in the process of measurement, there are at least three pairs of public feature points in the measurement process, the multi-view point cloud can be precisely registered. Through two coordinate transformations based on the positioning points, the registration of two point set can be achieved.

### III. ACCURACY ANALYSIS OF THREE-POINT POSITIONING METHOD

From the above data transformation method it can be seen that the alignment accuracy of model data depends on the measurement accuracy of three selected reference point. In addition, in the same measurement error circumstances, selection of different reference points will also affect the alignment model data. However if the error is controlled within certain range, such data transformation is able to meet the requirements of modelling and assembly.

Tao [12] proposed multiple measurement of datum points method which used two datum points and centroid as the new triangle to reduce registration error. To analyze the error of two transform method, define the vector difference of the three reference points as:

$$a_1 = P_2 - P_1, b_1 = P_3 - P_2, c_1 = P_1 - P_3;$$

$$a_2 = q_2 - q_1, b_2 = q_3 - q_2, c_2 = q_1 - q_3$$

When measurement errors exist, because the three non-collinear points determine a triangle, if we take the conversion method based on the three reference points, in fact, it is to ensure the overlap of a point and an edge. Fig. 2 shows the situation that  $P_1$  and  $q_1$  points overlap,  $a_1$

and  $a_2$  edges overlap and we define  $a_2 > a_1, c_1 > c_2$ .

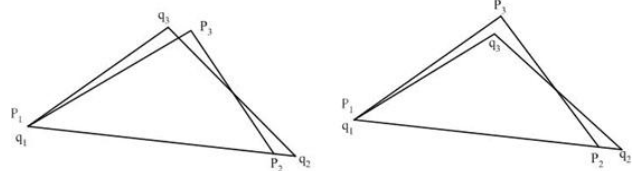


Figure 2. Three datum-points alignment model.

Define:

$$\varepsilon_1 = |a_1 - a_2|, \varepsilon_2 = |b_1 - b_2|, \varepsilon_3 = |c_1 - c_2| \quad (10)$$

Then the relative error can be expressed as:

$$\Delta_1 = \frac{|a_1 - a_2|}{|a_1|}, \Delta_2 = \frac{|b_1 - b_2|}{|b_1|}, \Delta_3 = \frac{|c_1 - c_2|}{|c_1|} \quad (11)$$

From equation (11), we can draw the following two conclusions:

(1) When the measurement error is constant, the bigger area of the triangle formed by the three points, then the smaller the relative error, that means the greater distance of reference points, the smaller impact of measuring errors on data alignment;

(2) In the case of normal distribution of measurement errors, the errors of three sides tend to be the same. The relative error should tend to be equal for the same impact of the various points. That is, the selection of reference point should be as close to an equilateral triangle.

#### IV. METHODS AND EXPERIMENT RESULTS

Since the error of each reference point can be seen as equal weight value, the relocation errors can be seen as average distributed errors. If we take a feature point of reference marker as the calibration reference point every time, the possibility of occurrence of human errors and accidental errors will increase greatly. Therefore, we can calculate the centroid of the vertices of reference marker and then use the centroid as the reference point to reduce registration errors. Specific methods are as follows:

Take equilateral triangle markers as artificial reference markers, there are three vertices of each reference point, shown in Fig. 3(a) and (b).

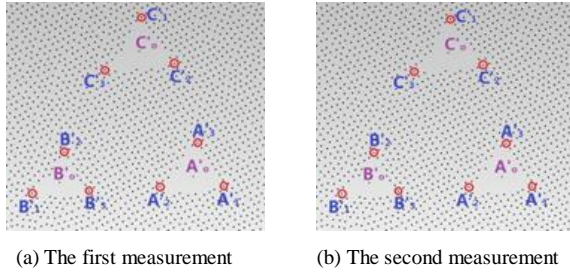


Figure 3. Three points alignment based on triangle reference markers.

First, take three vertices of reference markers  $A_1, B_1, C_1$  as benchmark reference points, the corresponding vertices are  $A'_1, B'_1, C'_1$ . The coordinates of each vertex is shown in Table 1:

TABLE I. VERTEX COORDINATES OF THE SELECTED REFERENCE MARKERS

Unit:mm	X Coordinate	Y Coordinate	Z Coordinate
$A_1$	63.751	52.445	925.525
$B_1$	26.602	-77.182	937.412
$C_1$	121.803	-56.026	958.979
$A'_1$	82.544	47.635	930.928
$B'_1$	60.090	-85.418	947.497
$C'_1$	152.957	-52.739	966.540

We can easily get the lengths of three sides of  $\triangle A_1 B_1 C_1$  and  $\triangle A'_1 B'_1 C'_1$  respectively, from equation (11), the relative errors can be expressed as:

$$\Delta_1 = \frac{\|B_1 C_1\| - \|B'_1 C'_1\|}{\|B_1 C_1\|} = 0.0041, \Delta_2 = 0.0014, \Delta_3 = 0.0043.$$

Then calculate the centroid coordinates of each vertex of triangle reference markers and use centroids as the new benchmark reference point, vertex coordinates and centroid coordinates are shown in Table 2.

TABLE II. THE COORDINATES OF THE VERTICES OF REFERENCE MARKERS AND THE CENTROIDS

Unit: mm	X Coordinate	Y Coordinate	Z Coordinate
$A_1$	63.751	52.445	925.525
$A_2$	62.908	45.588	926.450
$A_3$	68.714	48.663	927.720
$A_0$	65.125	48.899	926.565
$B_1$	26.602	-77.182	937.412
$B_2$	32.108	-75.148	938.562
$B_3$	27.180	-71.638	936.488
$B_0$	28.630	-74.656	937.487
$C_1$	121.803	-56.026	958.979
$C_2$	116.215	-52.159	956.649
$C_3$	115.439	-58.752	957.885
$C_0$	117.819	-55.646	957.838
$A'_1$	82.544	47.635	930.928
$A'_2$	82.639	41.045	932.060
$A'_3$	88.507	44.432	933.298
$A'_0$	84.563	44.370	932.096
$B'_1$	60.070	-85.418	947.497
$B'_2$	65.673	-82.440	948.553
$B'_3$	59.994	-79.176	946.404
$B'_0$	61.912	-82.345	947.485
$C'_1$	152.957	-52.739	966.540
$C'_2$	146.920	-49.158	964.127
$C'_3$	146.520	-56.128	965.577
$C'_0$	148.799	-52.675	956.415

It is easy to get the lengths of  $|A_0 B_0|, |B_0 C_0|, |C_0 A_0|$  and  $|A'_0 B'_0|, |B'_0 C'_0|, |C'_0 A'_0|$ , calculate their relative errors:

$$\Delta'_1 = \frac{\|B_0 C_0\| - \|B'_0 C'_0\|}{\|B_0 C_0\|} = 0.0027,$$

$$\Delta'_2 = 0.0010, \Delta'_3 = 0.0012.$$

Compare of the relative errors of two measurements, as shown in Table 3:

TABLE III. THE RELATIVE ERRORS OF TWO MEASUREMENTS

The first measurement		The second measurement	
$\Delta_1$	0.0041	$\Delta'_1$	0.0027
$\Delta_2$	0.0014	$\Delta'_2$	0.0010
$\Delta_3$	0.0043	$\Delta'_3$	0.0012

It can be seen that  $\Delta'_1 > \Delta_1, \Delta'_2 > \Delta_2, \Delta'_3 > \Delta_3$ , the precision increase greatly after we used the new triangle formed by centroids to substitute the original triangle and then the centroids can be used as reference points for

registration. Therefore, the three points coordinate transformation method can be improved to:

Step 1: Calculate the vertex coordinates of each reference triangle (polygon) markers;

Step 2: Calculate of the centroid coordinates of triangle (polygon) reference markers;

Step 3: Use centroids to form a new of triangle;

Step 4: Turn to the front algorithm step 1, replace the three measurement benchmark points by the new triangle centroids.

We apply the method to register scan data and reconstruct the electric vehicle shape plastic parts model from clay model in reverse engineering. Fig. 4 shows the electric vehicle front panel reconstruction model.

The first scanning data is chose as the stationary part, the other panel data is transformed to it.

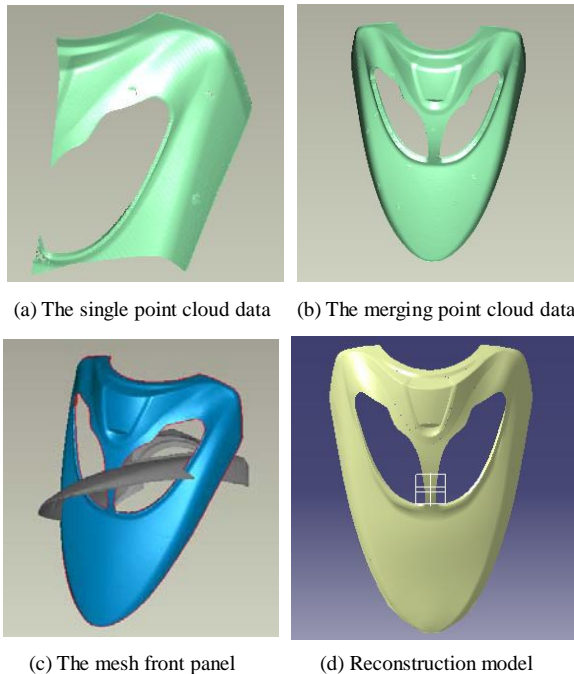


Figure 4. Electric vehicle front panel reconstruction model.

## V. CONCLUSION

Multi-view data alignment and the relocation is one of fundamental data processing problem in reverse engineering, a variety of methods has been proposed and 3-D point sets positioning method is a simple and practical

one among them. In this paper, we use the centroids instead of original vertices of reference markers to register points cloud data and apply the method to make scan data registration. Compare to other implementations the "centroid of apexes" method is more practical and time-saving. In many case, we can also use feature points on the surface of object directly instead of references markers. Experiment results show that the new method can quickly and effectively improve registration accuracy as well as being easy to use.

## REFERENCES

- [1] Ullrich, A., et al., Long-range high-performance time-of-flight-based 3D imaging sensors, in *3D Data Processing Visualization and Transmission 2002*: Padova, Italy. p. 852-855.
- [2] Zexiao, X., W. Jianguo, and J. Ming, Study on a full field of view laser scanning system. *International Journal of Machine Tools and Manufacture*, 2007. 47(1): p. 33-43.
- [3] Gorpas, D., K. Politopoulos, and D. Yova, A binocular machine vision system for three-dimensional surface measurement of small objects. *Computerized Medical Imaging and Graphics*, 2007. 31(8): p. 625-637.
- [4] Salvi, J., J. Pagès, and J. Battle, Pattern codification strategies in structured light systems. *Pattern Recognition*, 2004. 37(4): p. 827-849.
- [5] Li, L., et al., A reverse engineering system for rapid manufacturing of complex objects. *Robotics and Computer-Integrated Manufacturing*, 2002. 18(1): p. 53-67.
- [6] Larsson, S. and J.A.P. Kjellander, Motion control and data capturing for laser scanning with an industrial robot. *Robotics and Autonomous Systems*, 2006. 54(6): p. 453-460.
- [7] P.J. Besl, N.D. McKay, A Method for Registration of 3-D Shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2: p. 239-256.
- [8] Arun, K.S., T.S. Huang, and S.D. Blostein, Least-Squares Fitting of Two 3-D Point Sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1987. PAMI-9(5): p. 698-700.
- [9] Faugeras, O. D., and Hebert, M. The representation, recognition, and locating 3-D objects. *International Journal of Robotics Research*, 1986. 5(3): p. 27-52.
- [10] Horn, B. K. P. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 1987. 4(4): p. 629-642.
- [11] Mortenson, M.E., *Geometric modeling*, 2006: Industrial Press.
- [12] Tao, J. and K. Jiyong, A 3-D point sets registration method in reverse engineering. *Computers & Industrial Engineering*, 2007. 53(2): p. 270-276.

# Representing the Process of Machine Tool Calibration in First-order Logic

S. Parkinson, A. P. Longstaff, A. Crampton, S. Fletcher, G. Allen, A. Myers  
Centre for Precision Technologies, School of Computing & Engineering, University of Huddersfield,  
Queensgate, Huddersfield, HD1 3DH, UK  
simon.parkinson@hud.ac.uk

**Abstract**— Machine tool calibration requires a wide range of measurement techniques that can be carried out in many different sequences. Planning a machine tool calibration is typically performed by a subject expert with a great understanding of International standards and industrial best-practice guides. However, it is often the case that the planned sequence of measurements is not the optimal. Therefore, in an attempt to improve the process, intelligent computing methods can be designed for plan suggestion. As a starting point, this paper presents a way of converting expert knowledge into first-order logic that can be expressed in the PROLOG language. It then shows how queries can be executed against the logic to construct a knowledge-base of all the different measurements that can be performed during machine tool calibration.

**Keywords**—component; machine tool calibration; first-order logic; PROLOG

## I. INTRODUCTION

The continuing desire to manufacture artefacts to a higher degree of accuracy while decreasing the production cost has resulted in the requirement for more accurate machine tools [1-3]. This is because excess error within a machine tool's capability will manifest to the artefact during machining, possibly resulting in out-of-tolerance parts that are scrapped or require re-work. By calibrating a machine tool, the asset owner can gain an understanding of the machine's capability.

In a perfect world, the machine would be able to move to predictable points in 3-dimensional space, resulting in a machined artefact that is geometrically identical to that of the designed part. It is, however, well known that the machining process contains many possible sources of error that make it extremely unlikely for the ideal case to prevail. Machine tool pseudo-static errors can be classified into three general classes of; (1) rigid-body geometric errors, (2) thermally induced errors, and (3) non-rigid errors [4-6]. Precalibrated compensation is the process of measuring the machine to establish the pseudo-static geometric errors that will be transferred to the workpiece during machining to implement corrective action. In practice a machine tool has considerably more dynamic and thermal error components, however, for the scope of this paper, consideration is only given to the pseudo-static geometric errors for a machine tool with three perpendicular linear axes.

Great effort has been spent by many to improve the process of machine tool measurement to correctly identify the machine's error components. International standards [7] and best-practice guides [8] provide guidance regarding the selection of test methods for individual error

components. These are critical for performing meaningful measurements on machine tools. However, even with this rich knowledge there is still a great deal of interpretation, selection and planning to be done to develop a good strategy for measuring a specific machine tool [9]. For example, ISO 230-2 [7] contains a section regarding test specification parameters that need to be "agreed" between the calibration supplier, manufacturer and the user. This results in the need for "independent" expert knowledge because often the user does not have the sufficient level of experience to make this decision and sometimes the interests of the manufacturer and user can conflict, leaving an expert to make the best decision to suit all. Furthermore, downtime for calibration is a cost to manufacturing, so optimising the workflow has distinct commercial advantages.

Computational intelligence potentially holds the key to allow for a more efficient method of planning machine tool calibration. The use of Artificial Intelligence (AI) in the form of knowledge-based planning can allow for the efficient searching of a very large search space. This could be beneficial for the process of machine tool calibration if the knowledge and decision-making skills of an experienced machine tool metrology expert can be interpreted in the form of a computer program.

In this paper a literature survey of the developments in dimensional metrology process planning is presented, followed by a discussion on the process of machine tool calibration to establish a basic set of requirements. The method of separating the requirements into first-order logic for the PROLOG language leading to the creation of a knowledge-base is described, followed by how to calculate the estimated time for performing the sequence of tasks in the knowledge-base. Finally, the paper concludes by laying out future plans for the complete implementation of an intelligent machine tool calibration process planner.

## II. LITERATURE REVIEW

Within the subject of dimensional metrology, the effort required to create models and process guidelines are significant. However, it is common that the models are tailored for the measurement of specific, complex parts. Muelaner et al [10] presents a semi-autonomous method of metrology instrumentation selection for large-volume measurement based upon the artefact's dimensional criteria. This is performed by querying a database of known metrology instrumentation based upon the artefact's dimensions. However, the developed

method is heavily dependent on user interpretation. For example, if the identified artefact was to contain several complex dimensional characteristics, the user would be required to identify and include each one within the query, followed then by checking for any instrument-related physical access and visibility obstructions. As a result, this method can only be used to aid the suggestion of instrumentation as human verification is always required. As briefly suggested by the author, an algorithm that was to automatically determine the artefact's dimensional characteristics and identify any physical access and visibility issues would be highly beneficial. This would allow for the model to be used with an increased degree-of-confidence and significantly reduce the need for human verification, therefore, saving time. A modification of this method would not be suitable for machine tool calibration because a more intelligent solution is required which will always attempt to find the optimal sequence of measurements.

Planning in AI terms is reasoning about the effects of actions and the sequencing of available actions to achieve a given cumulative effect [11]. The use of AI planning techniques has been explored in many subject areas, including manufacturing [12]. Large efforts are spent developing AI within robot control and improving their level of reasoning [13]. The developments in these areas are significant; however their application is vastly different from that of machine tool calibration. There is an absence of any literature indicating that implementing AI techniques within machine tool calibration has previously been attempted, even though the potential gains are significant.

### III. CALIBRATION PLANNING

To be successful at planning, it is essential to first comprehensively understand the problem in hand, the desired achievement, and the method of arriving there. Following this principle, the following section gives an overview of machine tool calibration.

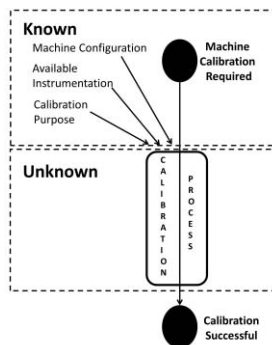


Figure 1. Known and unknowns sections of machine tool calibration

Figure 1 illustrates the known and unknown aspects of machine tool calibration. To overcome the unknown aspect, a machine tool metrology expert is responsible for deriving a “plan-of-action” for measuring the machine’s error components. To produce an autonomous method of planning, a method of defining the knowledge of a

machine tool calibration expert in first-order logic was discovered.

#### A. Expert’s knowledge

The knowledge and decision making procedure that a machine tool metrology expert possesses can be expressed as the following list of functions:

1. The ability to analyse the machine’s configuration to derive a set of tests.
2. The ability to determine the possible/best equipment for each test.
3. The ability to determine the best order of tests.

These three areas of decision making are vast, so to create an effective program to perform the required functionality resulted in a detailed investigation to identify the components that can be represented in a logical structure suitable for programming.

#### B. Machine Configuration

A machine can be constructed in many different ways to perform its task, and knowing the construction of the machine is essential for understanding the error components. These can be established because the geometric errors associated with both linear and rotary axes are known [1-5, 7, 14]. As shown in Table I, a linear axis will have six error components (six-degrees-of-freedom) plus a squareness error with the perpendicular axis. For the work presented in this paper, attention is only given to a machine tool with three perpendicular linear axes, which is illustrated in Figure 2.

Table I. Error Components

Linear Error Component
Linear positioning error
Horizontal straightness
Vertical straightness
Roll error
Pitch error
Yaw error
Squareness error

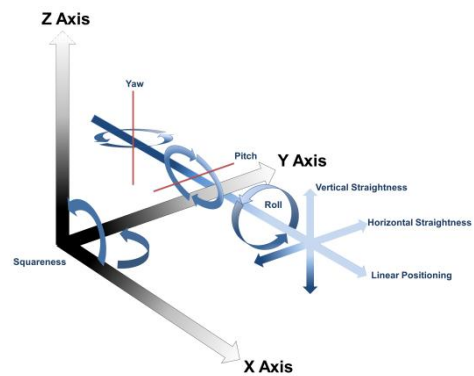


Figure 2. Six-degrees-of-freedom and squareness errors for the X-axis of a machine tool with three perpendicular linear axes



### C. Possible Tests

Once the array of error components has been derived based upon the machine's constituent parts and their configuration, a method of measuring the error component is required. This is where the expert's opinion takes precedence. There is no definitive method of measuring an error component, only suggestions provided by ISO 230 [7] and industrial best practice guides [8]. The following list gives a very simplistic example of a test method that could be used to measure each of the seven error components.

- 1) Linear positioning error - Laser interferometry to compare the machine's reported position against the actual position.
- 2) Horizontal straightness - Laser interferometry to map the horizontal deviation.
- 3) Vertical straightness - Granite straight edge and dial test indicator to map the vertical deviation.
- 4) Roll error - Electronic level to record the angular roll deviation.
- 5) Pitch error - Electronic level to record the angular pitch deviation.
- 6) Yaw error - Laser interferometry to record the angular yaw deviation.
- 7) Squareness error - Using a metrological square and a dial test indicator.

As a starting point for creating a prototype model, a common set of parameters to distinguish a measurement must be established. These parameters are as follows:

- 1) Axis under test - Axis name
- 2) Maximum axis travel - Length in mm
- 3) Number of measurements - Quantity
- 4) Number of repeats - Quantity
- 5) Direction of travel - Positive, negative, bidirectional.
- 6) Accuracy - Quantity in the test's unit
- 7) Resolution - Quantity in the test's unit
- 8) Out of the box setup time - Quantity of time in minutes
- 9) Setup time by readjustment - Quantity of time in minutes

### D. Best Order

As previously stated, the expert's execution of this method is different depending on their experience. Some individuals might prefer to perform the tests axis by axis, as some might prefer to perform the tests by the lowest effort of instrumentation setup. However, overall it is desirable for the most efficient method to be selected, which allows the calibration to be executed in the quickest time.

## IV. CALIBRATION SYSTEM PLANNER

### A. PROLOG

PROLOG is a programming language which is well suited to solving problems that involve objects and their relations [11, 15]. The work undertaken within this project could certainly be implemented in any procedural language; however, implementing symbolic computation by the use of PROLOG can considerably reduce the required quantity of code. For this reason, this work makes use of PROLOG language to produce a working prototype.

### B. Relation Facts

Representing the information for performing a machine calibration in relation facts allows for dynamic processing. The relation facts can be regarded as the facts for the knowledge engine which would allow the planner to make intelligent decisions. For the prototype program, the following facts were used. The facts show the relationship in first-order logic for a machine tool with three perpendicular linear axes, which each have seven geometric errors that use different instrumentation for measurement. These facts lack a great quantity of information that is required to perform a machine tool calibration, but as an example to develop a working prototype, they are sufficient.

```
%%% Machine type and axis name
machine(three_axis_machine, x_axis).
machine(three_axis_machine, y_axis).
machine(three_axis_machine, z_axis).
```

```
%%% Axis name, type and travel
axis(x_axis, linear, 500).
axis(y_axis, linear, 325).
axis(z_axis, linear, 490).
```

```
%%% Axis type and error component
error(linear, position).
error(linear, horizontal_straightness).
error(linear, vertical_straightness).
error(linear, roll)
error(linear, pitch).
error(linear, yaw).
error(linear, squareness).
```

```
%%% Error component and measurement method
test(position, laser_interferometer).
test(horizontal_straightness, laser_interferometer).
test(vertical_straightness, granite_straight_edge_dti).
test(roll, electronic_level).
test(pitch, electronic_level).
test(yaw, laser_interferometer).
test(squareness, metrological_square_dti).
```

```
%%% Measurement method, accuracy (µm), resolution
(µm), setup time and adjust time.
instrument(laser_interferometer, 1, 0.001, 30, 15).
instrument(metrological_square_dti, 1, 0.001, 20, 15).
instrument(granite_straight_edge_dti, 1, 0.001, 15, 8).
instrument(electronic_level, 1, 0.001, 18, 9).
```

Creating a PROLOG program with these facts allows us to essentially ask the program questions to retrieve useful knowledge. Consider the following prolog question;

Where: TY = Axis type, LEN = Axis travel length.

**?- axis(x\_axis, TY, LEN).**

Here a query is being executed to return the axis type and axis travel by asking the following question; what is the type and travel for the axis 'x\_axis'?

Executing this query would give us the response of;

**Type (TY) = linear,  
Travel (LEN) = 500.**

### C. Knowledge

In the same way as asking the PROLOG program simple questions, complex questions can also be asked to acquire knowledge. For planning a machine tool calibration, it would be beneficial to know all the properties of the specified three axis machine, including each axis type, all the geometric error components per axis and a method of measuring them. Dynamically creating a knowledge-base at run time can be achieved by creating a PROLOG procedure. The following procedure creates a list of all the possible tests to be performed and then prints them out for the user to see. See TABLE II for a key to the variables.

**knowledge\_base(M):-**

**setof( [A, TY, E, I, AC, RE, ST, SA] ,  
(machine(M, A), axis(A, TY, LEN),  
error(TY, E),  
test(T, I), instrument(I, AC, RE, ST, SA)),  
List), print\_kn(List).**

**print\_kn([H|T]):-**

**write(H), nl, print\_kn(T).**

The following command initiates the procedure;

**?- knowledge\_base(three\_axis\_machine).**

PROLOG would interpret the supplied facts and try to answer the question in as many ways as possible. This will provide the required information regarding each of the possible axis, error, test and instrumentation relationships. The response for this query contains 21 sets of information (7 tests per linear axis). However, adding additional instrumentation and error facts would result in an increased knowledge base. Executing this query against the previously established facts will result in the output that can be seen in TABLE II. The produced knowledge-base is a list of all possible tests. Currently there is no intelligence to disregard and order the tests respective to the machine's configuration and available instrumentation.

TABLE II. KNOWLEDGE BASE OF ALL AVAILABLE TESTS

Axis (A)	Type (TY)	Error (E)	Instrumentation (I)	Accuracy (AC)	Resolution (RE)	Setup Time (ST)	Adjust Time (AT)
x_axis	linear	position	laser_interferometer	1	0.001	30	15
x_axis	linear	horizontal straightness	laser_interferometer	1	0.001	30	15
x_axis	linear	vertical straightness	granite_straight_edge_dti	1	0.001	15	8
x_axis	linear	roll	electronic_level	1	0.001	18	9
x_axis	linear	pitch	electronic_level	1	0.001	18	9
x_axis	linear	yaw	laser_interferometer	1	0.001	30	15
x_axis	linear	squareness	metrological_square_dti	1	0.001	20	15
y_axis	linear	position	laser_interferometer	1	0.001	30	15
y_axis	linear	horizontal straightness	laser_interferometer	1	0.001	30	15
y_axis	linear	vertical straightness	granite_straight_edge_dti	1	0.001	15	8
y_axis	linear	roll	electronic_level	1	0.001	18	9
y_axis	linear	pitch	electronic_level	1	0.001	18	9
y_axis	linear	yaw	laser_interferometer	1	0.001	30	15
y_axis	linear	squareness	metrological_square_dti	1	0.001	20	15
z_axis	linear	position	laser_interferometer	1	0.001	30	15
z_axis	linear	horizontal straightness	laser_interferometer	1	0.001	30	15
z_axis	linear	vertical straightness	granite_straight_edge_dti	1	0.001	15	8
z_axis	linear	roll	electronic_level	1	0.001	18	9
z_axis	linear	pitch	electronic_level	1	0.001	18	9
z_axis	linear	yaw	laser_interferometer	1	0.001	30	15
z_axis	linear	squareness	metrological_square_dti	1	0.001	20	15



#### D. Simple planner

##### 1) Cost Calculator

A procedure which approximates overall cost in minutes for performing a machine tool calibration has been implemented. This procedure evaluates the potential cost of performing a measurement based upon the following six factors:

1. The time taken to set up the instrumentation out of the box.
2. The time taken to set up the instrumentation by readjustment.
3. The axis that the previous test was performed on.
4. The axis that the current test is being performed on.
5. The instrumentation used in the previous test.
6. The instrumentation used for the current test.

Implementing these six facts in PROLOG can be achieved with the following code:

Where: ST = Setup time, AT = Adjust time, PA = Previous axis, CA = Current axis, PI = Previous instrumentation, CI = Current instrumentation, COST = The overall cost in minute.

```
calccost(ST,AT,PA,CA,PI,CI,COST):-
  ( CA=PA ->
    (CI=PI ->COST=AT )
  ; COST=ST
  );COST=ST.
```

Adding this procedure to the simple planner results in the additional output of the cost for each measurement. The result from this execution can be seen in TABLE III.

#### V. CONCLUSION AND FUTURE WORK

The work undertaken within this paper has described how to represent a simplified version of performing machine tool calibration by representing the parameters as logical facts in PROLOG, and then reasoning about those facts. This was done by breaking the calibration requirements down into first-order logic regarding the machine's axes, error components and the available instrumentation. Next PROLOG's ability to answer complex questions that provide a knowledge-base for further processing is shown. This resulted in the implantation of a method for calculating the cost of performing a sequence of measurements from the knowledge-base. This will serve as a baseline to compare a future intelligent implementation.

The implemented method allows for the representation of an expert's knowledge that can subsequently be retrieved by anybody. However, because this method is based on simplified facts, it is not a complete solution that matches the real world scenario of machine tool calibration. The developed PROLOG program is a proof of concept which will lead to the implementation of a much more sophisticated and intelligent solution. The intelligent approach will require the creation of more facts and planning procedures to provide for a more complete application.

TABLE III. KNOWLEDGE OUTPUT INCLUDING CALCULATED COST

Cost	Axis	Type	Error Component	Instrumentation
30	x_axis	linear	position	laser_interferometer
15	x_axis	linear	horizontal_straightness	laser_interferometer
15	x_axis	linear	vertical_straightness	granite_stright_edge_dti
18	x_axis	linear	roll	electronic_level
9	x_axis	linear	pitch	electronic_level
30	x_axis	linear	yaw	laser_interferometer
20	x_axis	linear	squareness	metrological_square_dti
30	y_axis	linear	position	laser_interferometer
15	y_axis	linear	horizontal_straightness	laser_interferometer
15	y_axis	linear	vertical_straightness	granite_stright_edge_dti
18	y_axis	linear	roll	electronic_level
9	y_axis	linear	pitch	electronic_level
30	y_axis	linear	yaw	laser_interferometer
20	y_axis	linear	squareness	metrological_square_dti
30	z_axis	linear	position	laser_interferometer
15	z_axis	linear	horizontal_straightness	laser_interferometer
15	z_axis	linear	vertical_straightness	granite_stright_edge_dti
18	z_axis	linear	roll	electronic_level
9	z_axis	linear	pitch	electronic_level
30	z_axis	linear	yaw	laser_interferometer
20	z_axis	linear	squareness	metrological_square_dti
<b>Total:</b> <b>411</b>				

## VI. REFERENCES

- [1] J. Mou, "A systematic approach to enhance machine tool accuracy for precision manufacturing," *International Journal of Machine Tools and Manufacture*, vol. 37, pp. 669 - 685, 1997.
- [2] R. W. Bagshaw and S. T. Newman, "Manufacturing data analysis of machine tool errors within a contemporary small manufacturing enterprise," *International Journal of Machine Tools and Manufacture*, vol. 42, pp. 1065 - 1080, 2002.
- [3] R. Ramesh, *et al.*, "Error compensation in machine tools -- a review: Part I: geometric, cutting-force induced and fixture-dependent errors," *International Journal of Machine Tools and Manufacture*, vol. 40, pp. 1235 - 1256, 2000.
- [4] D. G. Ford, *et al.*, "The identification of non-rigid errors in a vertical machining center," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 213, pp. 555-566, 1999.
- [5] X. J. Wan, *et al.*, "A unified framework of error evaluation and adjustment in machining," *International Journal of Machine Tools and Manufacture*, vol. 48, pp. 1198 - 1210, 2008.
- [6] S. Parkinson, *et al.*, "A novel framework for establishing a machine tool quality metric," in *Future Technologies in Computing and Engineering Proceedings of Computing and Engineering Annual Researchers' Conference 2010 CEARC'10*, G. Lucas and Z. Xu, Eds., ed Huddersfield: University of Huddersfield, 2010.
- [7] ISO230, "Part 1: Geometric accuracy of machines operating under no-load or finishing conditions," in *Test code for machine tools*, ed. Geneva, Switzerland: International Standards Organisation, 1996.
- [8] NPL. (2005). *Measurement Good Practice Guide No. 80*.
- [9] S. Parkinson, *et al.*, "Controlling Machine Tool Accuracy Through a Robust Calibration Process," presented at the Yorkshire and North East Vitae Public Engagement Competition, Durham Town Hall, 2011.
- [10] J. E. Muelaner, *et al.*, "Large Volume Metrology Instrument Selection and Measurability Analysis," *IMechE*, 2010.
- [11] I. Bratko, *PROLOG Programming for Artificial Intelligence*, Third ed.: Pearson Education Limited, 2001.
- [12] J. Madejski, "Survey of the agent-based approach to intelligent manufacturing," *Journal of Achievements in Materials and Manufacturing Engineering*, vol. 21, pp. 67-70, 2007.
- [13] G. Lim and I. Suh, "Robust robot knowledge instantiation for intelligent service robots," *Intelligent Service Robotics*, vol. 3, pp. 115-123, 2010.
- [14] H. Schwenke, *et al.*, "Geometric error measurement and compensation of machines--An update," *CIRP Annals - Manufacturing Technology*, vol. 57, pp. 660 - 675, 2008.
- [15] K. A. Bowen, *Prolog and expert systems*. New York: McGraw-Hill, 1991.

# Localisation Algorithm in Wireless Sensor Networks

Shuang Gu<sup>1</sup>, Yong Yue<sup>1</sup>, Carsten Maple<sup>1</sup>, Chengdong Wu<sup>2</sup>

<sup>1</sup>Department of Computer Science and Technology, University of Bedfordshire, <sup>2</sup>School of Information Science and Engineering, Northeastern University

<sup>1</sup>Luton LU1 3JU, UK, <sup>2</sup>Shenyang, Liaoning, China

**Abstract**—This paper reviews different approaches to node localisation discovery in wireless sensor networks (WSNs) and addresses issues of localisation by describing a variety of localisation techniques and their strengths and drawbacks. The overview of the existing algorithms proposed for the improvement of localisation in WSNs is also presented. To understand how the performance of localisation techniques is evaluated, several evaluation metrics are described and the comparative results verified via Matlab simulations.

**Keywords**—localisation; Range-free; Simulation

## I. INTRODUCTION

Progress in micro-electromechanical systems (MEMS) and radio frequency (RF) technology has fostered the development of WSNs. WSNs have been successfully applied in emergency rescue, disaster relief, smart buildings and patient monitoring. As one of the key enabling technologies and research hotspots, node localisation is very important due to its direct correlation with theory and practical applications. Node localisation is required to report the origin of events, assist group querying of sensors, routing and to answer questions on the network coverage. [1][2][3]

Localisation algorithms can be classified to range-based and range-free algorithms [4][5]. In a range based algorithm, nodes determine their location based on distance or angle estimations to some reference points with known coordinates. In a range free localisation algorithm, nodes determine their location without using time, angle or power measurements, but use beacons, or connectivity information to compute their location.

The remainder of the paper is organised as follows. Section 2 discusses the state of the art work in localisation for WSNs. Section 3 describes the existing range-free localisation algorithm. Section 4 describes the simulation setting. Section 5 follows with a detailed performance comparison of the range-free localisation algorithms described. Finally we conclude in section 6.

## II. STATE OF THE ART

The Global Positioning System (GPS) [6] provides an immediate solution to the problem of localising a node in outdoor scenarios. However, having GPS in all the nodes is often not a viable solution in many scenarios due to its cost, the associated energy consumption and its inapplicability in different scenarios. Localisation

algorithms can be divided into two categories: range-based and range-free. There are many ranging methods for localisation which include Time of Arrival (TOA), Time Difference of Arrival (TDOA), Received Signal Strength (RSS), Near Field EM Ranging (NFER) and so on [7].

### A. Range-Based Localisation Algorithm

Range-based localisation algorithms depend on distance or angle between nodes to obtain unknown node's location. The first step is distance estimation or angle estimation [8]. A number of approaches, such as TOA, TDOA, AOA and RSS have been presented. The second step is location calculations. Trilateration, triangulation and maximum likelihood estimation are typical methods.

However, range-based algorithms are sensitive to environmental changes [9]. The problems that arise include (i) incorrect range measurements due to non line-of-sight conditions for acoustic ranging, (ii) error in RSS-based ranging caused by variations in channel parameters across different environments ( $1/r^n$  models), (iii) poor correlation between RSS and distance owing to multipath interference and fading, and (iv) insufficient number of reachable beacons or interference amongst densely deployed beacons for proximity-based localisation. These physical effects are difficult to predict and can lead to incorrect measurements which would greatly affect the quality of localisation in multilateration and iterative multilateration approaches.

### B. Range-Free Localisation Algorithm

Range-free localisation algorithms use the information of topology and connectivity for location estimation. Range-free algorithms have some advanced characteristics, such as low cost, small communication traffic, no extra hardware and flexible localisation precision. Because of these special characteristics of range-free algorithms, they were been regarded as a promising solution for the localisation problem in WSNs [10].

## III. RANGE-FREE ALGORITHM

### A. Centriod Localisation

Centroid localisation algorithm uses anchor beacons, containing location information, to estimate node position [11]. After receiving these beacons, a node estimates its location using the following centroid formula:

$$(X_{est}, Y_{est}) = \left( \frac{X_1 + \dots + X_k}{k}, \frac{Y_1 + \dots + Y_k}{k} \right) \quad (1)$$

where  $X_{est}$  and  $Y_{est}$  are the estimated coordinates of the sensor node and  $X_{lk}, Y_{lk}$  are the coordinates of all  $k$  anchor nodes heard by the sensor node.

### B. DV-Hop Localisation

DV-Hop localisation [12] is proposed by D. Niculescu and B. Nath in the Navigate project. DV-hop localisation uses a mechanism that is similar to classical distance vector routing. In this work, one anchor broadcasts a beacon to be flooded throughout the network containing the anchors location with a hop-count parameter initialised to one. Each receiving node maintains the minimum counter value per anchor of all beacons it receives and ignores those beacons with higher hop-count values. Beacons are flooded outward with hop-count values incremented at every intermediate hop. Through this mechanism, all nodes in the network (including other anchors) get the shortest distance, in hops, to every anchor.

In order to convert hop count into physical distance, the system estimates the average distance per hop without range-based techniques. Anchor performs this task by obtaining location and hop count information for all other anchors inside the network. The average single hop distance is then estimated by anchor using the following formula:

$$Hopsize_i = \frac{\sum_{j \neq i} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{j \neq i} h_j} \quad (2)$$

In this formula,  $(x_j, y_j)$  is the location of anchor  $j$ , and  $h_j$  is the distance, in hops, from anchor  $j$  to anchor  $i$ . Once calculated, anchors propagate the estimated HopSize information out to the nearby nodes.

Once a node can calculate the distance estimation to more than 3 anchors in the plane, it uses triangulation (multilateration) to estimate its location. Theoretically, if errors exist in the distance estimation, the more anchors a node can hear the more precise localisation will be.

### C. Amorphous Localisation

The Amorphous localisation algorithm, proposed independently from DV-Hop, uses a similar algorithm for estimating position. First, like DV-Hop, each node obtains the hop distance to distributed anchors through beacon propagation.

Once anchor estimates are collected, the hop distance estimation is obtained through local averaging. Each node collects neighbouring nodes hop distance estimates and computes an average of all its neighbours' values. Half of the radio range is then deducted from this average to compensate for error caused by low resolution.

The Amorphous localisation algorithm takes a different approach from the DV-Hop algorithm to estimate the average distance of a single hop. This work assumes that the density of the network  $n_{local}$  is known *a priori*, so that it can calculate HopSize offline in accordance with the Klein rock and Slivester formula:

$$Hopsize = r \left( 1 + e^{-n_{local}} - \int_{-1}^1 e^{-\frac{n_{local}}{\pi} (\arccos t - t\sqrt{1-t^2})} dt \right) \quad (3)$$

Finally, after obtaining the estimated distances to three anchors, triangulation is used to estimate a node's location.

### D. APIT Localisation

APIT requires a heterogeneous network of sensing devices where a small percentage of these devices (percentages vary depending on network and node density) are equipped with high-powered transmission and location information obtained via anchor nodes. Using beacons from these anchors, APIT employs an area-based approach to perform location estimation by isolating the environment into triangular regions between beaconing nodes. A node's presence inside or outside of these triangular regions allows a node to narrow down the area in which it can potentially reside. By utilizing combinations of anchor positions, the diameter of the estimated area in which a node resides can be reduced, to provide a good location estimate.

## IV. SIMULATION SETTING

### A. Transmission Model

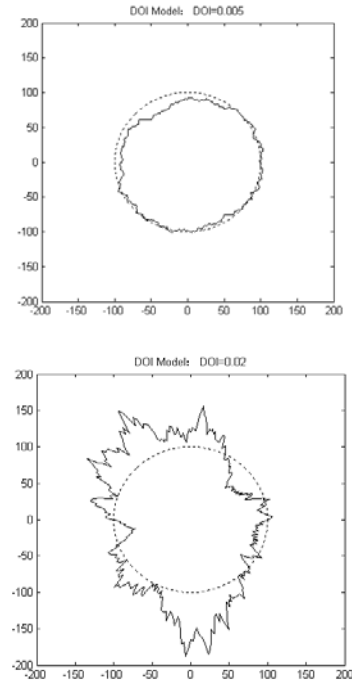


Figure 1. DOI model

- RIM model
- Regular model
- DOI model
- Logarithmic attenuation model

### B. Deployment Model

- Random deployment: it distributes all nodes and anchors randomly throughout the terrain. (e.g. C random)
- Uniform placement: the terrain is partitioned into grids and nodes and anchors are evenly divided amongst these grids (e.g. Square regular)

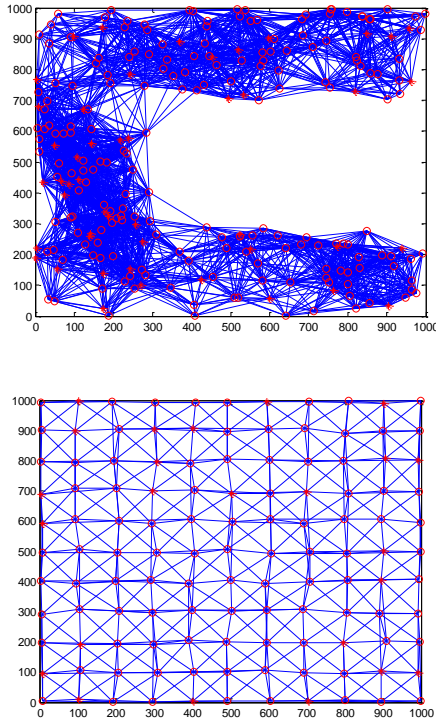


Figure 2. C random and Square regular deployment

### C. System Parameters

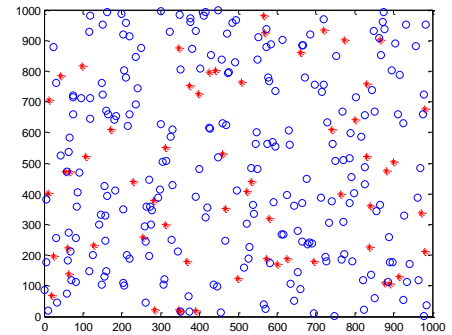
- Anchor Percentage (AP): The number of anchors divided by the total number of nodes.
- Degree of Irregularity (DOI): DOI is an indicator of radio pattern. It is defined as the maximum radio range variation per unit degree change in the direction of radio propagation.
- GPS Error: In reality, GPS equipped anchor will render imprecise readings. In this evaluation, it is defined as the maximum possible distance from the real anchor position to the GPS estimated anchor position in units of node radio range.
- Communication Radius (CR): The communication radius of anchor node is the proportion of the communication radius of unknown node.

- Deployment Error (DE): The difference between the grid point and the actual position of the sensor node.
- Localisation Error (LE): The average of localisation error is the Euclidean distance between the estimated location and actual location divided by communication radius.

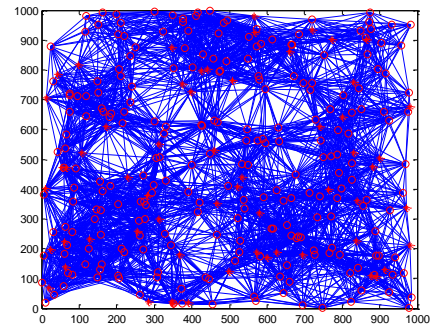
## V. EVALUATION

In the simulation, 4 different kinds of range-free localisation algorithms (DV-Hop localisation, Amorphous localisation, APIT localisation and Centriod localisation) are run on various topologies of networks in Matlab. Several system-wide parameters are studied that directly affect localisation error, connectivity and the number of neighbour anchor. In APIT, the sensor models are not homogeneous (grid length). The communication range of anchors is larger than that of unknown nodes. Communication radius represents the communication range of the unknown node and the communication range of the anchor node can be calculated by  $\text{times} \times \text{com\_r}$ .

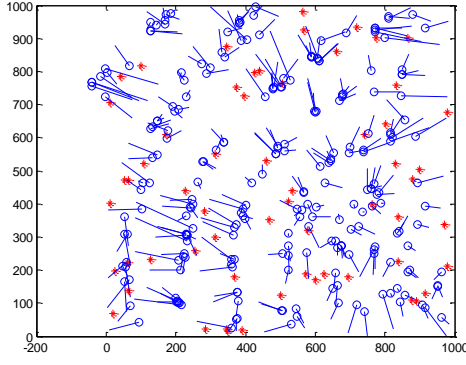
From Fig. 3 It can be seen that there are 300 sensor nodes and 60 anchor nodes (which the position is known before located). The red \*s represent anchor nodes and the blue 0s represent the unknown nodes. The blue —s represent the localisation error. The parameter is set as follows. The GPS error is 0. The communication model is regular model. The communication radius of anchor is 200 m.



(a)



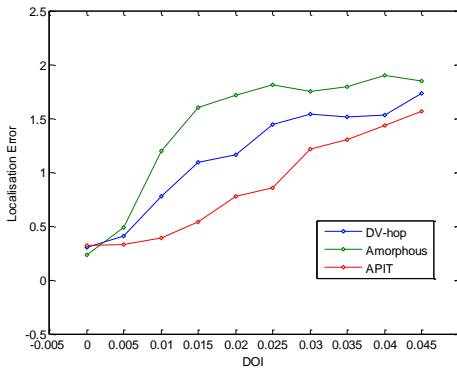
(b)



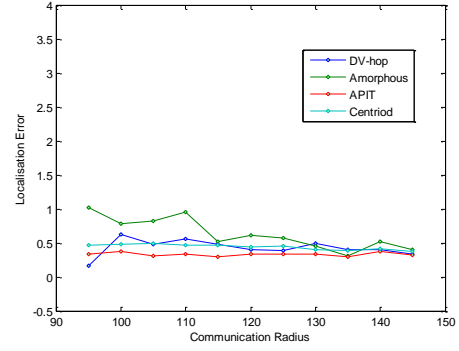
(c)

Figure 3. (a) Distribution of WSNs (b) Topology of WSNs (c) Localisation error

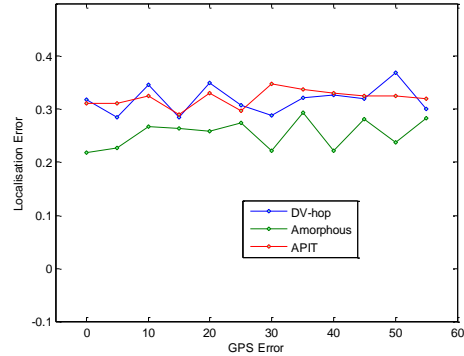
Simulation results are shown in Fig. 4 Fig.5 and Fig.6. In Fig.4a, DOI can affect the network topologies resulting in irregular hop count distributions in Amorphous and DV-hop algorithms. The HopSize formula, used in the Amorphous algorithm, assumes that radio patterns are perfectly circular. It can be also seen from the figure how this inaccurate estimate directly affects to localisation error as the DOI increases. The HopSize based algorithms can be easily affected by the anisotropic topologies (e.g. Fig. 5b C random model) or irregular transmission models (e.g. Fig. 5a logarithmic attenuation model). DV-hop algorithm estimates HopSize using online information exchanged between anchors. The results of DV-hop are much better performance than Amorphous (shown in Fig. 4a, Fig. 5a 5b 5c). Because the Centroid and APIT algorithms do not depend on hop-count and HopSize estimations, these algorithms seem to be good for better localisation when compared with HopSize based algorithm. However, they require a dense numbers of anchor nodes distributed over the whole network because direct communication between sensor nodes and anchors is required. Fig. 4c demonstrates how initial location error at anchors directly affects the error of the range-free localisation studied. In general, in all these algorithms GPS error is abated considerably by utilizing location information from multiple anchors.



(a)



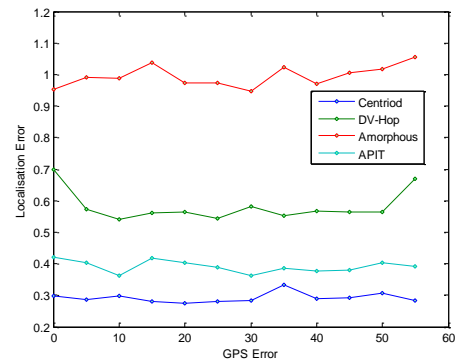
(b)



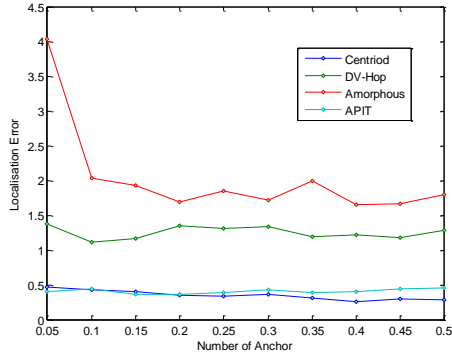
(c)

Figure 4. Localisation error when varying different parameters (DOI, number of anchor nodes, GPS error) in regular transmission model

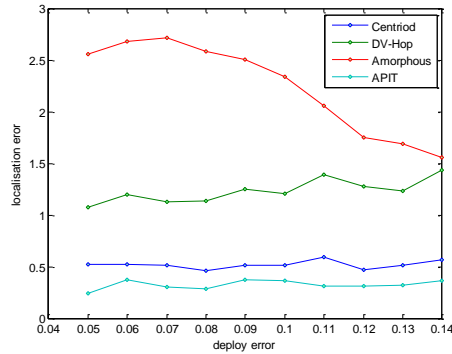
The effect of number of anchor on the localisation error is shown in Fig. 5d and the effect on average number of neighbour nodes is shown in Fig. 6b. As expected, the higher anchors density, the lower the localisation error and higher average number of neighbour nodes. The reason is that when there is higher number of anchors, each unknown node has more anchors that has multihop neighbourhood and gets more location information. The localisation algorithm is improved with less localisation error. The effect of communication radius on the localisation error is shown in Fig 4b, 6a. The effect of connectivity becomes high when the communication radius increases. The localisation error decreases due to increase in the communication radius.



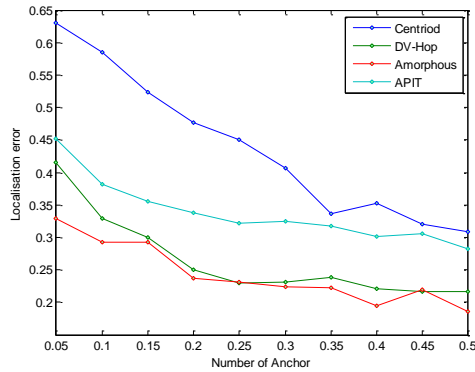
(a) Logarithmic attenuation model



(b) C random model

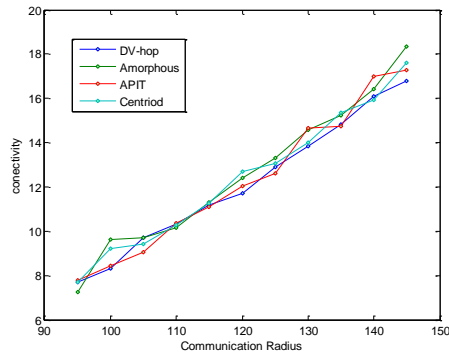


(c) C regular model

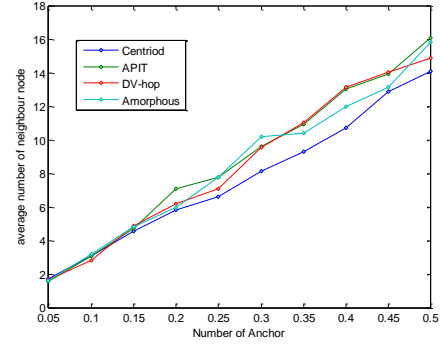


(d) Square regular model

Figure 5. Localisation error when varying different kinds of model



(a)



(b)

Figure 6. Connectivity and average number of neighbour nodes when varying parameters

## VI. CONCLUSION

In this paper, we have investigated 4 kinds of range-free localisation algorithm over different node deployments on different transmission models. The results show that the effect of each of the above parameters on any localisation algorithm depends on the actual techniques itself in terms of DOI or GPS error.

In the future, we will explore other localisation algorithms over a variety of realistic systems and designs protocols while considering different quantities for accurate, low-cost, low-energy purposes in a novel hybrid localisation algorithm in applications.

## REFERENCES

- [1] Zhetao Li et al. Survey of localisation techniques in wireless sensor networks. *Information technology Journal* 9 (8): 1754-1757, 2010. ISSN 1812-5638. 2010 Asian Network for Scientific Information. Pp: 1754-1757
- [2] Michal Marks and Ewa Niewiadomska-Szynkiewicz. Multiobjective Approach to Localisation in Wireless Sensor Networks. *Journal of telecommunications and information technology*. 2009.3. Pp: 59-66
- [3] Cao Xiaohong, Li Ying and Feng Huang. Research on Location of node for wireless sensor networks. *Information Technology*. 2009.7 Pp: 233-240
- [4] Isaac Amundson et al. RF angle of arrival-based node localisation. *Int. J. Sensor Networks*, Vol. 9, Nos. 3/4, 2011. Pp: 209-224
- [5] Frankie K. W. Chan, H. C. So, and W.-K. Ma. A Novel Subspace Approach for Wireless Sensor Network Positioning with Range Measurements. *ICASSP 2007* Pp 1037-1040
- [6] Nirupama Bulusu, John Heidemann and Deborah Estrin. GPS-less Low-Cost Outdoor Localisation for Very Small Devices. *IEEE Personal Communications* October 2000. Pp: 28-34
- [7] Oh-Heum Kwon, Ha-Joo song and Sangjoon Park. The Effects of Stitching Orders in Patch-and-Stitch WSN Localisation Algorithms. *IEEE transactions on parallel and distributed systems*, Vol. 20, No. 9, Pp1380-1391. September 2009
- [8] Kotwal, S.B.; Verma, S.; Suryansh, S.; Sharma, A.; Region Based Collaborative Angle of Arrival Localisation for Wireless Sensor Networks with Maximum Range Information. *Computational intelligence and communication networks (CICN)*, 2010 International Conference on 26-28 Nov. 2010 Pp: 301-307
- [9] Hyo-Sung Ahn and Wonpil Yu. Indoor Localisation Techniques based on Wireless Sensor Networks. Pp: 1-26
- [10] Amitangshu Pal. Localisation Algorithms in Wireless Sensor Networks: Current Approaches and Future Challenges. *Journal of Network Protocols and Algorithms* ISSN 1943-3581. 2010, Vol. 2, No.1 Pp: 45-74

- [11] Allon Rai, Sangita Ale and Syed S. Rizvi. A New Methodology for Self Localisation n Wireless Sensor Networks. Multitopic Conference, 2008. INMIC 2008. IEEE International Pp: 260-265
- [12] Dragos Niculescu and Badri Nath. DV based Positioning in Ad Hoc Networks. Telecommunication Systems 22: 1-4, 2003. Pp: 267-280



# User-defined gesture sets using a mobile device for people with communication difficulties

Yong Hee Jung<sup>1</sup>, Shengeng Qin<sup>2</sup>

School of Engineering and Design  
Brunel University

London, United Kingdom

<sup>1</sup>im10yhj@brunel.ac.uk, <sup>2</sup>Sheng.feng.qin@brunel.ac.uk

**Abstract**—Present smart phones contain high-tech sensors to monitor three-dimensional movements of the device and users' behaviours. These sensors allow mobile devices to recognize motion gestures. However, only a few gesture sets have been created, and little is known about best practices in motion-gesture design. Also, the created gesture sets were generated from people who do not have to use their motion gestures very much in their daily lives, not from people with communication difficulties. To address this issue, we use a focus group that has dyslexia and other specific learning difficulties for designing the user-defined gesture sets. This paper presents the results of our study that elicits the focus group's gestures to invoke commands on a smart-phone device. It demonstrates how the gesture sets have been designed and finalised throughout other research activities, such as observation and interviews. Finally, we suggest that our result would help people with communication impairments conveniently interact with others in an intuitive and socially acceptable manner.

**Keywords:** *communication; gesture; user interfaces; mobile devices*

## I. INTRODUCTION

Motion gestures play an important role in communication. So designing a set of good gestures using a mobile phone to educate or assist people with communication difficulties in an intuitive and socially acceptable manner is a prerequisite for the success of a gesture-based mobile communication system.

Gestures, the spontaneous hand or body movements that normally accompany face-to-face conversation [2], have been applied to devices using high-tech systems such as accelerometer-based motion-sensing techniques in mobile phones and camera-based methods. In particular, a gestural communication system using mobile phones has become a trend because present smart phones contain high-tech sensors to monitor three-dimensional movements of the device and users' behaviours. Thus, people highly expect that the high – tech systems will contribute to better life interactions for people who have communication difficulties. However, many researchers have studied the gestural interaction techniques to demonstrate how their gesture systems are more technically improved and unique and generally used in public, rather than for showing how the systems can be practically used for helping real users with communication difficulties [2].

Moreover, although a considerable number of gestural-communication systems with smart hand-held devices have been investigated since 1993 and introduced to public spaces, many people still suffer from communication difficulties in their daily lives. According to the Communication Forum, there are 2.5 million people in the UK with speech, language, and communication needs and more than 1 million children in the UK struggle to communicate [4]. Specifically, the largest group with communication disabilities in the UK is students in the 12 - 20 age range [9]. It has been shown that these young people develop emotional, social and behavioural difficulties, have lower academic achievement, and have mental issues, and these problems have led to unemployment and relationship difficulties after their studies. Given these facts, educating and supporting these people are important, and designing a specialised gesture-communication system that they can easily learn and use is paramount.

In fact, education researchers have shown that paying attention to gestures can be a powerful tool for training students and forming effective pedagogical responses. A growing body of research indicates that hand gestures play an important role in major subjects, such as art, languages, mathematics, and science. A number of studies report that students and teachers use gestures when they talk about science and teach about different languages [11]. Other research indicates that gesture shapes mathematical conversations in certain learning environments, often leading to new or more accurate task understanding [2]. In addition, researchers have used spontaneous hand gestures to assess children's problem-solving strategies in language learning and math [11].

However, a set of gestures, using gestural recognition techniques, has not been designed yet for educational contexts, and existing general gestures are not sufficiently useful for students. Also, most related studies using hand-held motion gestures have simply determined one method without considering the communication disabilities.

To solve these problems, this paper focuses specifically on gestures for young people with dyslexia and other learning difficulties to easily learn and use in the academic context. We specifically observed a focus group to find common gesture patterns in a library. Based on our observation, motion gestures for a variety of input tasks using different methods, such as navigating different hand-gesture methods and inputting text, are designed and introduced to make their future brighter and provide

sustainable development for people with communication impairments via a specialised gesture-based communication system.

## II. LITERATURE REVIEW

### A. Definition of Gesture

Several definitions of gesture exist. The most common use of the term is with respect to natural gestures, which are defined as supports to verbal communication [5]. For Cassel and colleagues, “A natural gesture means the types of gestures spontaneously generated by a person telling a story, speaking in public, or holding a conversation [5].” Offering a wider perspective, Poggi [2] defines gesture as “any movement performed by hands, arms, or shoulders.” We use gestures to do things, to touch objects, people, or oneself, and finally to communicate.

Under the wider-perspective definition of a gesture, a communicative gesture can be defined as when a particular form and movement of hands is used for the goal of communicating some meaning, that is, to help some other person understand some belief or idea [11]. A communicative gesture is thus a signal-meaning pair: The signal is a particular form and movement of the hands or arms; the meaning is a belief mentally represented either as a mental image or in a propositional format; and the signal and the meaning are linked to each other in the minds of some people [11].

Hands can do three kinds of things: rest, without doing anything; move to do some noncommunicative action; or, communicate. Thus we can distinguish:

- 1) Resting positions
- 2) Action gestures
- 3) Communicative gestures.

How to assess whether a gesture is either a communicative or an action gesture — that is, if it is ruled by the goal of communicating or simply by the goal of doing an instrumental action — is an empirical issue, which has to be assessed case by case on the basis of context of use.

### B. Previous Work in Gestural Recognition Systems

Gesture recognition is a wide field of different interaction techniques and devices that have the common goal of interpreting human bodily motions into computer input for interaction [3]. For effective gesture-based communications, the key issue is how gestural input devices interpret accurately a series of motions from users. So, many researchers have explored the gestural recognition techniques using diverse devices and sensors over the past 20 years.

Gesture systems were introduced in 1987 when glove-based user interfaces, called “DataGloves,” were created for interaction with wearable computers [16]. They were equipped with a number of sensors which provide information about hand position, orientation, and flex of the fingers. The glove measures each joint bend to an accuracy of 5 to 10 degrees [16]. Another gesture system was the Styli interface with display technologies to record and interpret gestures like the writing of text [17]. Finger-

based sensors detect finger positions and some tablet PCs work with electromagnetic-resonance pens. Position Trackers use ultrasound emissions and infrared light to identify the movements that make up a gesture [17]. Changes in ultrasound waves can measure the changes in a finger’s position relative to a fixed point.

After these developments, gesture recognition systems have become more practical. Systems are beginning to combine image- and device-based techniques to gather more information about gestures and researchers are also upgrading the sensors that relay information about users’ movements. Thus the upgraded systems enable more accurate recognition. In 2003, iMatte introduced iSkia, a technology that enables presenters to interact with projectors and screens using gesture recognition [17]. When presenters hold down buttons on a remote control, the iSkia system recognizes the movements of their extended hand and converts them into on-screen drawing or highlighting [17]. Recently, various types of gesture-recognition products, such as Microsoft’s Surface, a table with a touch-sensitive top responding to hand gestures and real-world objects, g-speak, special sensor gloves for detecting spatial hand gestures, and Ubiquitous window, which enables gestural interaction with a screen behind glass through optical motion detection, have been released in the market place [3, 7, 9]. However, these applications rely on expensive custom hardware and are not mobile.

Therefore, research in the field of mobile computing has also investigated the usage of acceleration sensors to detect manual gestures. One of the favoured uses is the interaction with connected real world objects such as distant displays through a gesture-aware mobile phone. Another possibility for detecting a phone gesture is to analyze the built-in camera’s video stream. With these enhancements in mobile hardware, more complex computer vision algorithms can be realized on smart phones leading to handheld augmented reality applications.

### C. Existing Smart Hand-Held Devices

A different set of gesture methods have been introduced with smart hand-held devices.

1) *Designing DoubleFlip*: According to Ruiz and Li [13], DoubleFlip is a unique motion gesture that acts as a delimiter for other motion gestures, and hence, separates normal mobile-phone motion from a user’s intended input. In addition, the gesture gives users the control to activate motion gestures without any hardware modifications to existing devices. The gesture is quick to perform. Since the gesture can be carried out with the simple rotation of the wrist and no movement in the arm, the gesture requires no more physical space than holding the phone for normal use [13].

Myers and Rabiner [14] assert that the ability to distinguish the DoubleFlip gesture from normal motion creates many possibilities for motion-gesture based interaction. It allows the user to activate and deactivate a motion gesture mode by simple flipping, and enlarges the design space of motion gestures by not requiring gestures to be different from ambient motion. The gesture scenario is that when another person asks you a simple question,

you can simply answer. For example, "Do you like apples?" The answer can be expressed by signs – Yes: Double flip toward the right, No: Double flip toward the left.

2) *Shoogle on Mobile Phones*: Shoogle is a novel, intuitive interface for sensing data, such as the presence and properties of text messages or remaining resources, within a mobile device [15]. It is based around active exploration: Devices are shaken, revealing the contents rattling around "inside." In the study by Williamson, Murray-Smith, and Hughs [15], a version of Shoogle has also been implemented in standard mobile-phone hardware (Nokia 6660 and other modern series 60 phones). It uses the custom-built Bluetooth SHAKE (Sensing Hardware Accessory for Kinesthetic Expression) Inertial sensor pack for sensing. Mobile phones that already incorporate accelerometers would require no additional hardware [1, 4, 6, 15].

Communications are over a Bluetooth serial port profile. SHAKE includes a powerful DSP engine, allowing a real-time linear-phase sample-rate conversion [15]. The SHAKE sensor allows realistic inertial sensing prototypes to be rapidly implemented on mobile phones, with plausible form factors and all of the accumulated data a real phone carries [4]. Its gesture scenario allows the user to reach into his or her bag, and shake the phone gently, without removing or looking at it. The contents of the SMS inbox are transformed into virtual "message balls." As the user shakes the phone, impacts are heard and felt as these balls bounce around. Hearing that one of them is distinctly metallic, and has a deep, heavy feeling impact, the user realises that a long text message from a colleague has been received.

3) *MagiWrite*: The basic idea is to draw gestures similar to characters (in terms of shape) in the 3-D space around the device [10]. The character-shaped gestures are made using a properly shaped magnet taken in hand. Some mobile devices such as iPhone 3GS and Google Android are equipped with a magnetic sensor, which is originally used for navigation purposes [1].

Movement of a magnet in a shape similar to a certain character can change the temporal pattern of the magnetic field around the device, and can be sensed and registered using the internally embedded magnetic sensor [10]. The idea is partly inspired by the Around Device Interaction (ADI) framework, which proposes using space around the device for interaction with the device [21]. ADI techniques are based on using different sensory inputs such as a camera, infrared distance sensors, a touch screen at the back of device, a proximity sensor, and electric field sensing [20]. The gesture scenario is when a person writes "2" around a mobile device, the device recognises the number movement, and the number shows up in the mobile screen.

4) *Whack Gestures*: This method seeks to provide a simple means to interact with devices with minimal attention from the user – in particular, without the use of

fine motor skills or detailed visual attention [12]. For mobile devices, Hudson, Harrison, and LaMarca [18] state that this could enable interaction without "getting it out," grasping or even glancing at the device. Thus, common interactions could be carried out in an extremely lightweight fashion without disrupting other activities.

The gestures are designed by a small vocabulary of gestures intended to interact with a small mobile device worn at the waist in a holder attached to a belt. Other locations, such as in the pocket or in the bag or purse, are also possible, and would likely only marginally affect the quality of the accelerometer-driven data we use for gesture recognition [12]. Gestures are performed by firmly striking the device with an open palm or heel of the hand moved towards the waist – an action referred to simply as a whack [8]. One of its gesture scenarios is that a person moves around his or her waist if the signal from a mobile phone on the waist is right.

All these systems are useful and have unique technologies. These methods have also been proven with diverse experiments for expressing users' thoughts and interacting with others. However, there is a drawback to each system when it is used individually as it is only beneficial to a limited group of people. For example, DoubleFlip and Shoogle cannot be used for people with mobile impairments, and MagiWrite is not useful for the blind because the blind cannot check whether or not a mobile phone recognises his or her gesture signs properly.

### III. METHOD

#### A. Research with Focus Group

This research was conducted at Brunel University, London, UK and took place in the University library, where students often use gestures to have quiet conversations. To set user-defined gestures, we specified one group whose members are dyslexic and have specific learning difficulties, such as problems with hearing, speaking, reading, and cognitive functions (see Table I).

TABLE I. DISABILITY CATEGORIES IN BRUNEL UNIVERSITY

Disability Categories in Brunel University	Main Problems					
	Listening	Speaking	Seeing	One hand	Both hand	Cognitive functions
Sensory Impairments						
Unseen Disabilities						
Mobility Impairments						
Dyslexia and Other specific learning difficulties						
Mental Health						

Disability and Dyslexia Service, Brunel University.

The group of students include three third-year undergraduate students — two females and one male.

Lynn (female) and Tanvi (female) are psychology students, and Josef (male) is a sports-science student. This group was selected for analysis with the goal of having diversity in the study in terms of both academic background gestures and gender gestures.

Observation of the focus group’s special and common actions was conducted in the Brunel University library, over four weeks for a total of twenty hours. We tracked their common gesture actions while they were doing each task in the library.

#### B. Selection of Task

From the observation, main tasks — finding a book, printing, and reading a book — are often done by students in the library. While conducting those tasks, the focus group showed common actions. The focus group performed actions for each of the tasks indicated in Table II.

TABLE II. MAIN TASKS IN THE LIBRARY

Main Tasks	Actions
Finding a book	1) Open library search engine
	2) Type the name of a book or author in the new search engine
	3) Confirm if input data is correct
	4) Delete previous text when the input data is incorrect
	5) Close the search engine
Printing	1) Turn on computer
	2) Open a file
	3) Set up printing options
	4) Click “OK” button
	5) Go to printer
	6) Tap student cards on the swipe-card device
	7) Select files that have already been loaded
	8) Push the button, “Print”
Reading a book	1) Take a book from a shelf
	2) Put the book on a desk
	3) Sit down in a chair
	4) Read the book

Of the main tasks, one task, finding a book via a library search engine, was selected to design a set of gestures that fits an academic context. This task was the most performed task in the library, and the focus group and general students expressed via interviews that they want to use simple and efficient gestures for the task.

#### C. Procedure

At the beginning of designing a set of gestures, diverse gesture recognition as a ubiquitous input for the mobile phone was selected. To effectively communicate with

someone or something, unlike with the existing devices which only used a single-gesture system, diverse system-gestures for each listed action were considered in this study. For example, the first step in the task, “Open,” could have different gesture sets when using a mobile phone, such as touching an iconic “Open” button on the phone, or moving two hands toward the outside while holding a mobile phone, or inputting text “Open,” or drawing the initial word “o” around the device to recognise it as the “open” command.

After designing the set of motion gestures for the given task set, in order to reaffirm our selected gestures, we asked the focus group to perform each gesture 10 times on cue and then rate each gesture using numbering from the minimum 1 to the maximum 5 on following criteria; ease of use, frequency of use, and fit for its intended use.

Furthermore, an interview session was conducted to evaluate the usefulness of the determined gestures. The focus group was asked how often each person would use the motion gesture if the gesture existed and if they had suggestions for other gestures that would be more beneficial.

#### IV. RESULTS

The data collected during our study included observation records with pictures and video, a set of gestures based on the given task “finding a book” from the observation, subjective ratings of the set of gestures, and focus-group suggestions in the interviews.

From the data, we finalised for each given task the fine gesture set that had the highest total scores.

##### A. Designing Set of Gestures Based on the Given Task

To be properly defined user gestures, all potential gestures were classified with each step of actions in the given task. The potential gesture sets for each step are designed and listed in Table III.

##### B. Subjective Ratings of the Designed Gesture Set

After we designed gestures for a particular task, the focus group rated the fit and ease of use, and how often they would use the gesture assuming it existed. The rating results are also in Table III.

TABLE III. POTENTIAL GESTURE SETS AND EVALUATION

Main Tasks	Actions	Potential Gesture Sets	Evaluation			
			Fit	Ease of use	Freq uency	Total
Finding a book	1) Open library search engine	• Touch an iconic “Open” button on the phone	4	4	4	12
		• Move two hands toward the outside while holding a mobile phone	3	3	2	8
		• Input text “Open”	2	2	2	6
		• Draw the initial word, “o” around the device	3	3	4	10
		• Snap the phone to the right once	4	5	4	13

Main Tasks	Actions	Potential Gesture Sets	Evaluation			
			Fit	Ease of use	Frequency	Total
	2) Type the name of a book or author in the new search engine	•Input the first and last words of the full name	3	2	1	7
		•Touch the initial word of the full name in order	4	4	3	11
3) Confirm if input data is correct		•Make a thumb up for "Yes" sign and down for "No" sign with holding a mobile phone	5	4	5	14
		•Tap the phone once for "Yes" and twice for "No."	3	5	4	12
		•Draw "O" around the phone for "Yes" and "X" for "No"	3	3	3	9
4) Delete previous text when the input data is incorrect		•Wipe their hand out to the right on the mobile screen	5	4	4	13
		•Shake the phone from the left to the right once	3	4	3	10
		•Touch an iconic "Delete" button on the phone	5	4	3	12
5) Close the search engine		•Snap the mobile phone to the left twice	4	5	4	13
		•Move two hands toward the inside	3	3	2	8
		•Touch an iconic "Close" button on the phone	4	4	4	12

### C. Finalised a User-Defined Gesture Set

After rating each gesture set from the focus group, final preferred-gesture sets for the given task were suggested. From the final gesture set in Table IV, we can see gesture that is a result of input from the focus-group.

The gesture for the first action, "Open library search engine," is to snap the phone to the right once. When the focus group tried to find a book via an online library search engine in the library, it was really hard for them to find the menu on the library website due to their reading problems. So, by making the mobile recognise a snap motion, they can command the action.

The gesture for the second action, "Type the name of the book or author in the new search engine" was to touch the initial word of the full name in order. The focus group members did not want to type all the letters of each word and could not read the whole lists of books. So, by simply touching each initial letter of the names of books or authors, they expect to minimize their time to find a book and not display their disabilities in public.

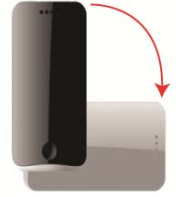




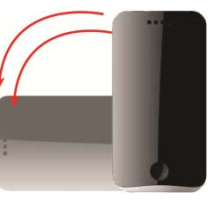
The gesture for the third action, "Confirm if input data is correct," is to make a thumb up for "Yes" and down for "No" while holding a mobile phone. Although there are many ways to express "Yes" or "No" signs, such as tapping the phone once for "Yes" and twice for "No," and drawing "O" around the phone for "Yes" and "X" for "No," the focus group preferred the ease of movement of simply using their thumbs to move up and down. Additionally, they said tapping the phone once or twice to express the signs would make them confused

sometimes because it may conflict with signs of numbers, like tapping the phone once = Number 1 and twice = Number 2.

When the input data is incorrect, another action gesture is created for deleting the previous text. The focus group chose the gesture, "Wipe the hand out to the right on the mobile screen," and thought this gesture was similar to the usual behaviour for erasing a whiteboard in a classroom. So, the gesture was easy to remember and perform.

After finishing all the actions for the task, the users needed to close the search engine. The focus group wanted to use the gesture that contrasted with the open action for the library search engine for easier memory and performance. Thus, the gesture for closing the search engine was to snap the mobile phone to the left twice.

TABLE IV. FINAL GESTURES FOR EACH ACTION

Main Tasks	Actions	Final Gestures	
Finding a book	1) Open library search engine	Snap the phone to the right once	
	2) Type the name of a book or author in the new search engine	Touch the initial word of the full name in order	<p>Example) The name of the book: Gesture-Based Communication in Human-Computer Interaction Author: Annelies Braffort</p>  <p>Touch initial word in order : G, B, C, H, C</p>
	3) Confirm if input data is correct	Make a thumb up for "Yes" sign and down for "No" sign with holding a mobile phone	  <p>Yes No</p>
	4) Delete previous text when the input data is incorrect	Wipe the hand out to the right on the mobile screen	
	5) Close the search engine	Snap the mobile phone to the left twice	

## V. FURTHER WORK

A limitation of our study is that the user-defined gestures were not tested with a specific mobile application. A part of the gesture set that was already defined could be improved because of actual test results. In future, we would develop a mobile application system to generate the gestures and test how the gesture set is properly working for the focus group.

In addition, our focus group was from one university. It is quite possible that gestures are influenced by culture. Thus, we will keep exploring the possible use of online tools that other university students with similar problems can revise, and we will expand the user-defined gesture set as the tasks that users wish to accomplish on mobile devices change. Moreover, there would be a test for general users on how the user-defined gestures could be used to address real-life communication.

## VI. CONCLUSION

In this paper, we described the results of a preliminary study of motion gestures through a specific focus group, students with dyslexia and other specific learning difficulties in an academic context. We showed how a detailed subset of gestures has been designed with particular user agreement and evaluation. As a result of a finalized gesture set, we presented design heuristics that inform motion-gesture design for mobile interaction. Finally, we highlighted the significance of gesture design based on a particular user group and gesture structures for an effective gestural interaction.

## REFERENCES

- [1] A. Butler, S. Izadi, and S. Hodges, "SideSight: Multi-touch interaction around small devices," in *UIST'08*, Monterey, California, USA, pp. 201-204, October 2008.
- [2] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," in *5<sup>th</sup> International Gesture Workshop*, Genova, Italy, pp.20-39, April 2003.
- [3] A. Savvas, M. Konstantinos, K. Alexey, A. Oya, T. Dimitrios, T. Thanos, V. Giovanna, and K. Byungjun, "Multimodal user interface for the communication of the disabled," *Journal of Multimodal User Interfaces*, vol. 2, pp. 105-116, June 2008.
- [4] A. Scoditti, R. Blanch, and J. Coutaz, "A novel taxonomy for gestural interaction techniques based on accelerometers," in *IUI'11*, Palo Alto, California, USA, pp. 63-72, February 2011.
- [5] B. Bossard, A. Braffort, and M. Jardino, "Some issues in sign language processing," in *5<sup>th</sup> International Gesture Workshop*, Genova, Italy, pp.90-100, April 2003.
- [6] E. Jones, J. Alexander, A. Andreou, P. Irani, and S. Subramnian, "Ges text: Accelerometer-based gestural text-entry systems," in *CHI 2010*, Atlanta, Georgia, USA, pp. 2173-2182, April 2010.
- [7] F. Anthony and A.M. Sorin, "Enhancing human-computer interaction design education: teaching affordance design for emerging mobile devices," *International Journal of Technology Design Education*, vol. 20, pp.239-254, January 2009.
- [8] G. Niezen and G. Hancke, "Gesture recognition as ubiquitous input for mobile phones," in *UbiComp'08 Workshop W1*, September 2008.
- [9] G. Rajat, G. Vikrant, and A. Vineet, "A smart mobility solution for physically challenged," in *Thirteenth IEEE International Symposium Wearable Computers (ISWC'09)*, pp. 168-173, 2009.
- [10] H. Ketabdard, M. Roshandel, "MagiWrite: Towards touchless digit entry using 3D space around mobile devices," in *MobileHCI'10*, Lisboa, Portugal, pp. 443-446, September 2010.
- [11] J. L. Hanna, "A nonverbal language for imagining and learning," *American Educational Research Association*, pp. 492-506, March 2008.
- [12] J. MIN, B. Choe, and S. Cho, "A selective template matching algorithm for short and intuitive gesture UI of accelerometer-builtin mobile phones," *Proceedings of the World Congress on Nature and Biologically Inspired Computing*, pp.667- 672, 2010.
- [13] J. Ruiz and Y. Li, "DoubleFlip: A motion gesture delimiter for mobile interaction," in *UIST'10*, New York, New York, USA, pp. 449-450, October 2010.
- [14] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," in *CHI 2010*, Vancouver, BC, Canada, May 2011.
- [15] J. Williamson, R. M. Smith, and S. Hughes, "Shoogle: Excitatory multimodal interaction on mobile devices," in *CHI 2007*, San Jose, California, USA, pp. 121-124, April 2007.
- [16] K. Holger, M. Friedrich, and S. Robert, "A glove-based gesture interface for wearable computing applications," in *Proceedings of the IFAWC 4<sup>th</sup> international forum on applied wearable computing 2007*, pp. 169-177, 2007.
- [17] S. Dan, *Designing Gestural Interfaces*, Sebastopol, CA: Cambridge, 2009.
- [18] S. E. Hudson, C. Harrison, and B. L. Harrison, "Whack gestures: Inexact and inattentive interaction with mobile devices," *Cambridge, Massachusetts, USA*, vol. 978-1, pp. 109-112, January 2010.
- [19] S. Jeon, G. J. Kim, M. Billingham, "Interaction techniques in large display environments using hand-held devices," in *VRST'06*, Limassol, Cyprus, November 2006.
- [20] S. Kallio, J. Kela, J. Mantyjarvi, and J. Plomp, "Visualization of hand gestures for pervasive computing environments," in *AVI'06*, Venezia, Italy, pp. 480-482, May 2006.
- [21] S. Ronkainen, J. Hakkila, S. Kaleva, A. Colley, and Jukka Linjama, "Tap input as an embedded interaction method for mobile devices," in *TEI'07*, Baton Rouge, Louisiana, USA, pp.263-270, February 2007.

# The Effect of Depth of Cut on the Molecular Dynamics (MD) Simulation of Multi-Pass Nanometric Machining

A.O. Oluwajobi<sup>1</sup> and X. Chen<sup>2</sup>  
Centre for Precision Technologies  
University of Huddersfield  
Queensgate, Huddersfield HD1 3DH, UK  
<sup>1</sup>j.o.oluwajobi@hud.ac.uk, <sup>2</sup>x.chen@hud.ac.uk

**Abstract—** The effect of depth of cut on multi-pass nanometric machining of copper workpiece with diamond tool was studied using the Molecular Dynamics (MD) simulation. The copper-copper interactions were modelled by the EAM potential and the copper-diamond interactions were modelled by the Morse potential. The diamond tool was modelled as a deformable body and the Tersoff potential was applied for the carbon-carbon interactions. It was observed that the average tangential and normal components of the cutting forces increase with increase in depth of cut and they reduced in consecutive cutting passes for each depth of cut. Also, the ratio of the tangential to normal force components decreases as the depth of cut increases, but remains fairly constant after 1.5nm depth of cut. The ratio of the cutting force to area decreases with increase in the depth of cut and remains constant after 2.5nm depth of cut.

**Keywords-** Multi-Pass; Depth of Cut, Molecular Dynamics; Nanometric Machining, Cutting Forces

## I. INTRODUCTION

Current material removal technological requirements in the aerospace, automobile, medical and energy industries are at the nanoscale, with stringent form and surface finish accuracy. At this length scale, machining phenomena take place in a small limited region of tool – workpiece interface, which often contains a few atoms or layers of atoms. At present, it is very difficult to observe the diverse microscopic physical phenomena occurring through experiments at the nanoscale [1]. The interface at this nanometre level may not be considered as a continuous media or homogeneous as assumed by continuum mechanics, so the analysis should be based on discrete atoms, whose interactions are governed by appropriate interatomic potentials. The use of Molecular Dynamics (MD) simulation has proved to be an effective tool for the investigation of machining processes at the nanometre scale. The method gives higher resolution of the cutting process than what is possible by continuum mechanics on that length scale [2].

The MD method was initiated in the late 1950s at Lawrence Radiation Laboratory in the US by Alder and Wainwright in the study of statistical mechanics [3].

Since then, the use of the simulation method has extended from Physics to Materials Science and now to Mechanical Engineering. Rentsch and Inasaki [4] modelled a copper workpiece and a diamond tool using the Lennard-Jones potential for the copper atom interactions. They observed a build-up phenomenon after 25000 time steps, while keeping the tool rigid. Komanduri et al [5] used copper workpiece and an infinitely hard tungsten tool for their simulation. They used Morse potentials and a cutting speed of 500m/s.

Many existing MD simulation studies on nanometric cutting have been limited to single pass or simple line-type groove. As an extension of the single pass studies, Zhang et al [6] modelled folder- line grooves for AFM-based nanometric cutting process of copper workpiece with diamond tool. They used the EAM potential for the copper-copper interactions and the Morse potential for the copper-diamond interactions. They treated the diamond tool as rigid and concluded that the normal, lateral and the resultant forces were almost symmetric with respect to the critical folder angle of 45°. Shi et al [7] investigated the multi-groove simulation of single-point turning of copper workpiece with diamond tool. They used two diamond tools, offset by a fixed distance to simulate a two-groove cutting and modelled the copper-copper and the copper-diamond interactions by using the Morse potential. They also treated the tool as a rigid body and observed that the tool forces increase with increase in feed rate and depth of cut. In practice, most machining processes involve the use of multiple passes to create new surface patterns and the diamond tool is deformable and subjected to wear. This study clearly shows the consecutive passes of cut, which is novel in multi-pass nanometric machining MD simulations. Also, the effect of the variation of depth of cut on the simulation of multi-pass cutting was investigated to model the surface creation in single point diamond turning.

## II. THE MD METHODOLOGY

The nanometric cutting model consists of a monocrystalline copper workpiece and a diamond tool.



The model configuration has a total of 54232 atoms as shown in Fig. 1. The workpiece is made up of 43240 copper atoms, with the face-centred cubic (FCC) lattice. It includes 3 kinds of atoms namely; boundary atoms, Newtonian atoms and thermostat atoms (See Fig. 1). The cutting tool has 10992 carbon atoms with diamond lattice structure. Fig. 2a shows a diagram of the machined grooves with passes 1 - 3 and Fig. 2b shows the tool tip dimensions, with the upper part as variable, which depends on the depth of cut considered.

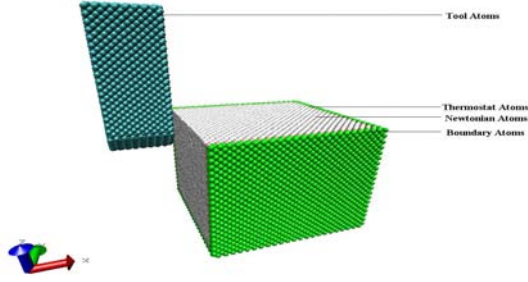


Figure 1: The MD Simulation Model

The end of the cutting tool is trapezoidal shaped (fairly pointed, with a blunt end). For the workpiece, the boundary atoms are kept fixed to reduce edge effects.

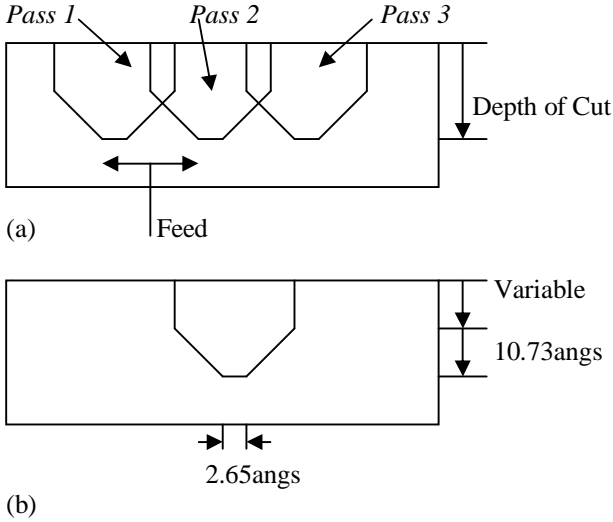


Figure 2a: Cross Section of the Machined Grooves with Passes 1-3 (direction of cut is perpendicular to the paper face) 2b: Tool Tip Dimensions

The Newtonian atoms obey the Newton's equation of motion. The thermostat atoms conduct the heat generated during the cutting process out of the cutting region. This is achieved by the velocity scaling of the thermostat atoms, (with the conversion between the kinetic energy (KE) and temperature via Eq. 1 [8, 9]);

$$\sum_i \frac{1}{2} m_i v_i^2 = \frac{3}{2} N k_B T_i \quad (1)$$

Where  $m_i$  is the mass of the  $i$ th atom,  $v_i$  is the resultant velocity of the  $i$ th atom,  $N$  is the number of the thermostat atoms,  $T_i$  is the temperature of the  $i$ th atom and  $k_B$  is the Boltzmann constant ( $1.3806504 \times 10^{-23} \text{ JK}^{-1}$ )

Whenever the temperature of the thermostat atoms exceeds the preset bulk temperature of 293K, their velocities are scaled by using Eq. 2 [10, 11];

$$v_{i,new} = v_i \sqrt{\frac{T_{desired}}{T_{current}}} \quad (2)$$

Where  $v_{i,new}$  is the newly scaled velocity of atom  $i$ ,  $v_i$  is the velocity of atom  $i$ ,  $T_{current}$  is the current temperature that is calculated from the KE and the  $T_{desired}$  is the desired temperature.

The simulation conditions applied in this study are the following, viz; bulk temperature is 293K, the cutting direction is along the x-axis, the cutting speed is 150m/s, the feed is 1.5nm, the time step is 0.3fs and the simulation run is 150000 steps. The depths of cut used are 0.5nm, 1.0nm, 1.5nm, 2nm, 2.5nm and 3 nm. The LAMMPS parallel MD software [12] was used for the simulations, on the University of Huddersfield's High Performance Computer (HPC) clusters, with a total of 144 processing cores (36 nodes) and a RAM of 8 GB (800Mhz) per node. The MD software utilized 2 nodes and 4 processors for each simulation. The VMD software [13] was used for the visualization of the results.

### III. THE MODELLING PARAMETERS FOR THE SIMULATION

It has been previously established that the EAM potential is very suitable for the Cu-Cu interactions [14], [15] and for the Cu-C interactions; the Morse potential is a good choice [16].

#### A. Embedded-Atom Method Potential (EAM) (Eq. 3) [17] (For the Cu-Cu interactions)

$$E_{tot} = \sum_i G_i(\rho_{h,i}) + \frac{1}{2} \sum_{i,j} V_{ij}(r_{ij}) \quad (3)$$

Where  $E_{tot}$  is the total embedding energy,  $\rho_{h,i}$  is the total electron density at atom  $i$  due to the rest of the atoms in the system,  $G_i$  is the embedding energy for placing an atom into the electron density,  $V_{i,j}$  is the short range pair interaction representing the core-core repulsion,  $r_{ij}$  is the separation of atoms  $i$  and  $j$ .



B. *Morse Potential* (Eq. 4) [18]  
(For the Cu-C interactions)

$$V_{ij} = D\{\exp[-2\alpha(r_{ij} - r_e)] - 2\exp[-\alpha(r_{ij} - r_e)]\} \quad (4)$$

Where  $V_{ij}$  is the pair potential,  $r_{ij}$  and  $r_e$  are instantaneous and equilibrium distances between atoms  $i$  and  $j$  respectively,  $\alpha$  and  $D$  are constants determined on the basis of the physical properties of the material.

The parameters used in the simulations are below, [19];

$$D = 0.087 \text{ eV}, \alpha = 0.17(\text{nm})^{-1}, r_e = 0.22 \text{ nm}$$

The cut-off distance chosen was 6.4 Angstroms (that is, the interactions between atoms separated by more than this distance are neglected).

B. *Tersoff Potential* (Eq. 5) [20]  
(For the C-C interactions)

$$E = \sum_i E_i = \frac{1}{2} \sum_i \sum_{j \neq i} V_{ij} \quad (5)$$

and,

$$V_{ij} = f_c(r_{ij})[a_{ij}f_R(r_{ij}) + b_{ij}f_A(r_{ij})]$$

where

$$f_R(r) = A \exp(-\lambda_1 r),$$

$$f_A(r) = -B \exp(-\lambda_2 r),$$

$$f_c(r) = \begin{cases} 1, & r < R - D \\ \frac{1}{2} - \frac{1}{2} \sin\left[\frac{\pi}{2}(r - R)/D\right], & R - D < r < R + D \\ 0, & r > R + D \end{cases}$$

$$b_{ij} = (1 + \beta^n \zeta_{ij}^n)^{-1/2n},$$

$$\zeta_{ij} = \sum_{k(\neq i, j)} f_c(r_{ik}) g(\theta_{ijk}) \exp[\lambda_3^3 (r_{ij} - r_{ik})^3],$$

$$g(\theta) = 1 + \frac{p^2}{q^2} - \frac{p^2}{[q^2 + (h - \cos \theta)^2]},$$

$$a_{ij} = (1 + \alpha^n \eta_{ij}^n)^{-1/2n},$$

$$\eta_{ij} = \sum_{k(\neq i, j)} f_c(r_{ik}) \exp[\lambda_3^3 (r_{ij} - r_{ik})^3]$$

Where  $E, E_i$  are the energies of interacting atoms,  $V_{ij}$  is the pair potential,  $R$  and  $D$  are cut-off parameters;  $A, B, \lambda_1, \lambda_2, \lambda_3, \alpha, \beta, n, p, q, h$  are fitting parameters of the Tersoff potential.

The simulation parameters used for carbon, are given as [20, 21, 22];

$$A(\text{eV}) = 1.3936 \times 10^3; B(\text{eV}) = 3.467 \times 10^2;$$

$$\lambda_1(\text{nm}^{-1}) = 34.879; \lambda_2(\text{nm}^{-1}) = 22.119; \alpha = 0.0;$$

$$\beta = 1.5724 \times 10^{-7}; n = 7.2751 \times 10^{-1}; p = 3.8049 \times 10^4;$$

$$q = 4.384; h = -5.7058 \times 10^{-1};$$

$$\lambda_3(\text{nm}^{-1}) = 22.119; R(\text{nm}) = 0.18; D(\text{nm}) = 0.02.$$

( $F_x, F_y, F_z$  : are the tangential, lateral and normal components of the cutting forces respectively (eV/Angs =  $1.602 \times 10^{-9}$  N)).

It should be noted that, generally, the interatomic potential parameters used for MD simulations are verified by ascertaining good agreement between their predicted values of the material properties and experimental data.

#### IV. SIMULATION RESULTS

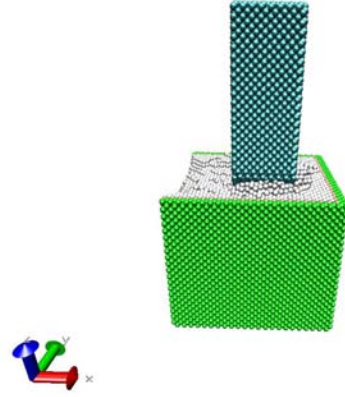


Figure 3: Simulation of Depth of Cut 0.5nm – Pass 3

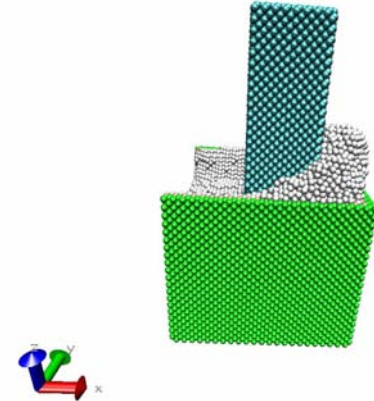


Figure 4: Simulation of Depth of Cut 1.5nm – Pass 3

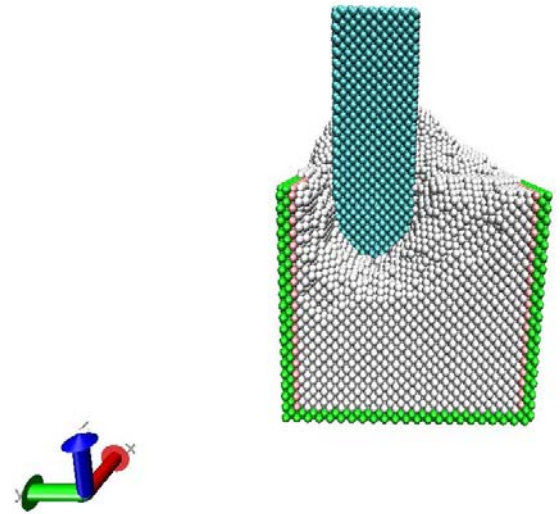


Figure 5: Simulation of Depth of Cut 3nm – Pass 1

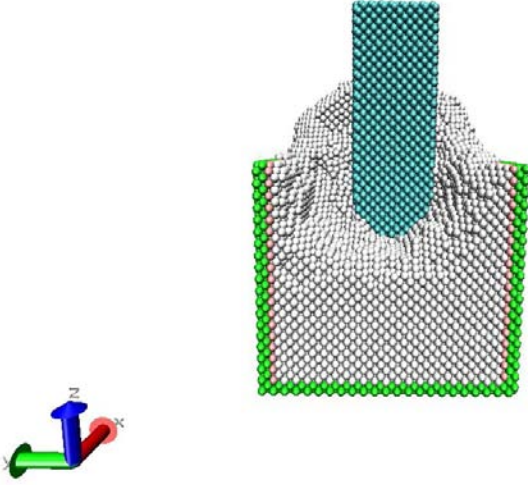


Figure 6: Simulation of Depth of Cut 3nm – Pass 2

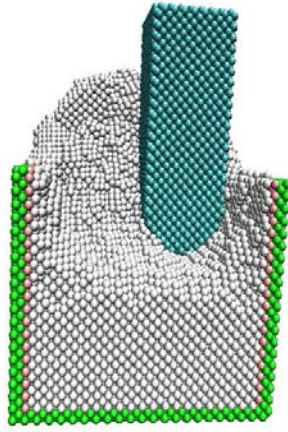


Figure 7: Simulation of Depth of Cut 3nm – Pass 3

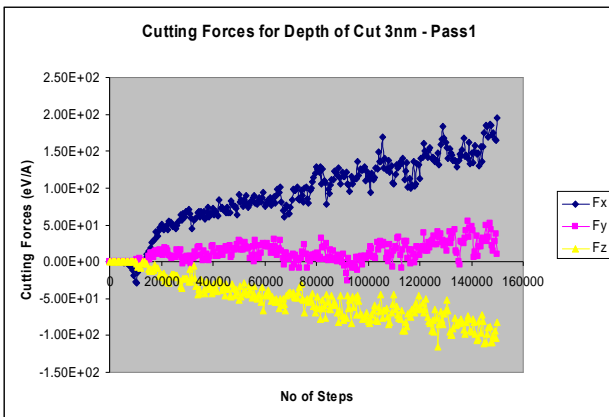


Figure 8: Cutting Forces for Depth of Cut 3nm – Pass 1

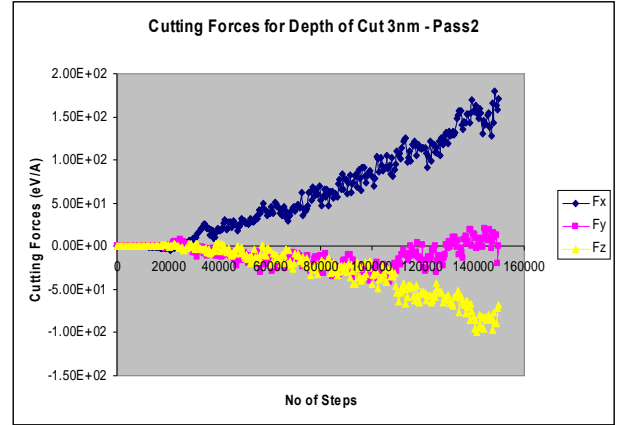


Figure 9: Cutting Forces for Depth of Cut 3nm – Pass 2

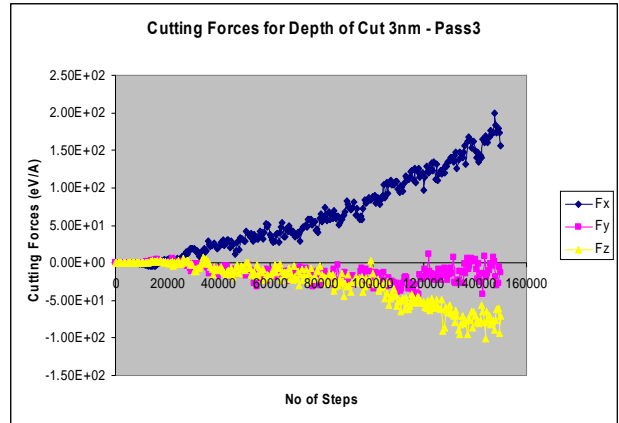


Figure 10: Cutting Forces for Depth of Cut 3nm – Pass 3

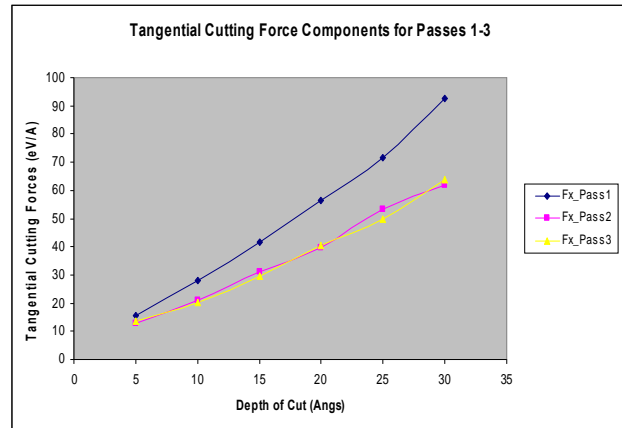


Figure 11: Variation of  $F_x$  in Passes 1-3 with Depth of Cut

## V. RESULTS AND DISCUSSION

Figs. 3 and 4 show the simulations after the third pass for the depth of cut of 0.5nm and 1.5nm respectively. From the figures, it can be observed that the amount of atoms removed increases as the depth of cut increases, which is logical, because as the depth increases, there is more volume of material atoms to be removed.

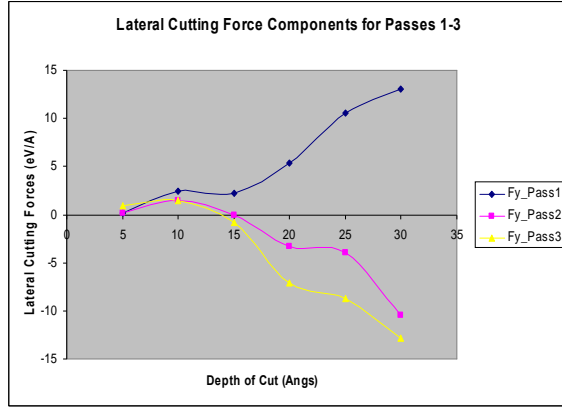


Figure 12: Variation of  $F_y$  in Passes 1-3 with Depth of Cut

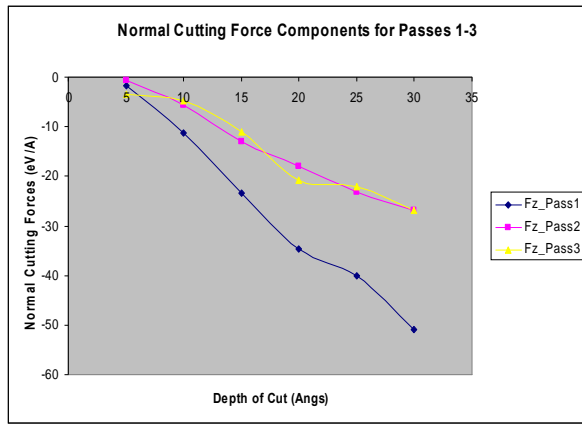


Figure 13: Variation of  $F_z$  in Passes 1-3 with Depth of Cut

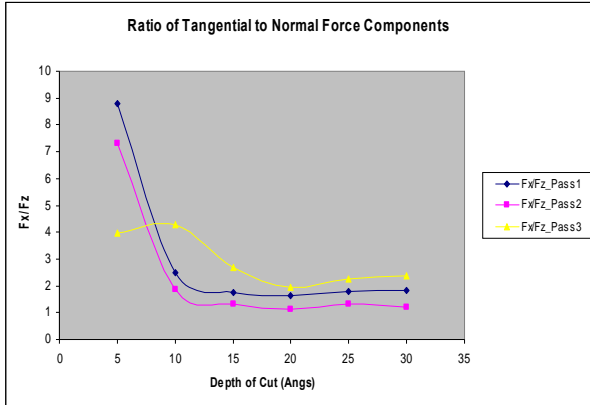


Figure 14: Variation of  $F_x/F_z$  in Passes 1-3 with Depth of Cut

Figs. 5-7 show the simulation of the three consecutive passes for the depth of cut of 3nm and Figs. 8-10 show the associated cutting forces respectively. The average tangential and the normal cutting force components decrease with the consecutive passes. Figs. 11-13 show the variation of the cutting force components, ( $F_x$ ,  $F_y$  and  $F_z$ ) with depth of cut, for passes 1-3. It can be seen that for the different passes, both  $F_x$  and  $F_z$  increase in magnitude with increase in the depth of cut. The average values of  $F_x$  and  $F_z$  are larger in pass 1 than in passes 2

and 3. Also, the values for passes 2 and 3 are quite close. This is because the cutting cross sectional area for passes 2 and 3 are similar. The variation of  $F_y$  is considerably smaller than both  $F_x$  and  $F_z$ , and it should be zero theoretically. The small variation is due to the atomic vibration during the cutting process. For pass 1,  $F_y$  is positive and becomes negative for passes 2 and 3 after the depth of cut of 1.5nm.

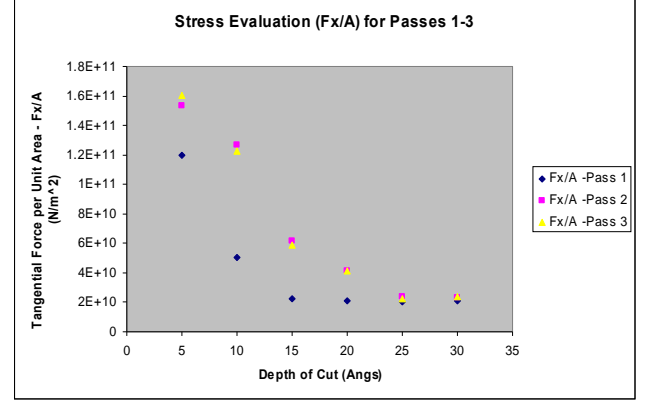


Figure 15: Stress Variation with Depth of Cut for Passes 1-3

The change from positive to negative is because for passes 2 and 3, the structure of the workpiece to be cut becomes asymmetrical and skewed to the right. Fig. 14 shows the ratio  $F_x/F_z$ , which is a measure of friction for the different passes. It can be seen that this ratio is fairly constant after the depth of cut of 1.5nm, which might be due to the tool geometry. Fig. 15 shows the stress variation with depth of cut for the three passes. It can be observed that as the depth of cut increases, the stress values decrease and it is higher for passes 2 and 3. The stress values are in the range from 160GPa to 20GPa. The values remain constant at around 20 GPa for all passes for higher depth of cut – from 2.5nm. This is due to the tool geometry, which becomes similar for higher depths of cut. The highest stress values are for depth of cut of 0.5nm during passes 2 and 3. It shows that the cutting resistance of the copper material is highest at very small depth cuts.

## VI. CONCLUSION

The simulation of multi-pass nanometric machining has been conducted by using the MD method, and the effect of varying the depth of cut has been investigated. Some important results are hereby outlined. The magnitude of the tangential and the normal components of the cutting forces increase with the increase in the depth of cut. The ratios of the tangential to normal force components decrease as the depth of cut increases, but remain fairly constant for each of the passes after 1.5nm depth of cut, with values in the range of 1.1-2.3. Stress values decrease with increase in the depth of cut and remain constant for high depth of cut.

## REFERENCES

- [1] Rentsch, R., "Nanoscale cutting", in Davim, J.P. and Jackson, M.J. (Eds.): *Nano and Micromachining*, Wiley-ISTE, 2008, pp. 1–24
- [2] Komanduri R. and L.M. Raff, "A Review on the Molecular Dynamics Simulation of Machining at the Atomic Scale", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 215, No. 12, 2001, pp. 1639-1672
- [3] Alder B.J. and T.E. Wainwright, "Studies in Molecular Dynamics. I. General Method", *Journal of Chemical Physics*, Vol 31, 1959, pp. 459-466
- [4] Rentsch R. and I. Inasaki, "Molecular Dynamics Simulation for Abrasive Processes", *Annals of the CIRP* Vol. 43, No 1, 1994, pp. 327-330
- [5] Komanduri R., N. Chandrasekaran and L.M. Raff, "Some Aspects of Machining with Negative-Rake Tools Simulating Grinding: A Molecular Dynamics Simulation Approach", *Philosophical Magazine Part B*, Vol. 79, No 7, 1999, pp. 955-968
- [6] Zhang J.T. , T. Sun, Y. Yan, Y. Liang and S. Dong, "Molecular Dynamics Study of Groove Fabrication Process using AFM-Based Nanometric Cutting Technique", *Applied Physics A (Materials Science and Processing)*, Vol. 94, 2009, pp. 593-600
- [7] Shi J., Y. Shi and C.R. Liu, "Evaluation of a Three-Dimensional Single-Point Turning at Atomistic Level by a Molecular Dynamics Simulation", *International Journal of Advanced Manufacturing Technology*, Vol. 45, Nos 1-4, 2011, pp. 161-171
- [8] Cai M., X. Li and M. Rahman, "Molecular Dynamics Modelling and Simulation of Nanoscale Ductile Cutting of Silicon", *International Journal of Computer Applications in Technology*, Vol. 28, No. 1, 2007, pp 2-8
- [9] Guo Y., Y. Liang, M. Chen, Q. Bai and L. Lu, "Molecular Dynamics Simulations of Thermal Effects in Nanometric Cutting Process", *Science China Technological Sciences*, Vol. 53, No. 3, 2010, pp. 870-874
- [10] Cheong W.C.D., L. Zhang and H. Tanaka, "Some Essentials of Simulating Nano-Surface Processes using the Molecular Dynamics Method", *Key Engineering Materials*, Vol. 196, 2001, pp. 31-42
- [11] Lin Z.-C. Z.-D. Chen and J.-C. Huang, "Establishment of a Cutting Force Model and Study of the Stress-Strain Distribution in Nano-scale Copper Material Orthogonal Cutting", *International Journal of Advanced Manufacturing Technology*, Vol. 33, No. 5-6 2007, pp. 425-435
- [12] Plimpton S. J., "Fast Parallel Algorithms for Short-Range Molecular Dynamics", *J Comp Phys*, Vol. 117, 1995, pp. 1-19 and [www.lammps.sandia.gov](http://www.lammps.sandia.gov)
- [13] Visual Molecular Dynamics (VMD), <http://www.ks.uiuc.edu/Research/vmd/> (Accessed in 2010)
- [14] Oluwajobi A.O. and X. Chen, "The Effect of Interatomic Potentials on Nanometric Abrasive Machining", *Proceedings of the 16<sup>th</sup> International Conference on Automation and Computing*, 2010, pp. 130-135
- [15] Oluwajobi A.O. and X. Chen, "The Fundamentals of Modelling Abrasive Machining Using Molecular Dynamics", *International Journal of Abrasive Technology*, Vol. 3, No. 4. 2010, pp. 354-381
- [16] Pei Q.X., C. Lu, F.Z. Fang and H. Wu, 'Nanometric Cutting of Copper: A Molecular Dynamics Study', *Comp.Mat. Sci.*, Vol. 37, 2006, pp. 434-441
- [17] Foiles S.M., "Application of the Embedded Atom Method to Liquid Transition Metals", *Physical Review B*, Vol. 32 No 6, 1985, pp. 3409-3415
- [18] Morse P.M., "Diatomic Molecules according to Wave Mechanics II Vibrational Levels", *Physical Review* Vol. 34, 1929, pp. 57-64
- [19] Hwang H.J., O-K Kwon and J. W. Kang, "Copper Nanocluster Diffusion in Carbon Nanotube", *Solid St. Comm.* 129, 2004, pp. 687-690
- [20] Tersoff J., "Empirical Interatomic Potential for Silicon with Improved Elastic Properties", *Physical Review B*, Vol. 38 No 14, 1988, pp. 9902-9905
- [21] Raffi-Tabar H. and G.A. Mansoori, "Interatomic Potential Models for Nanostructures", in *Encyclopedia of Nanoscience and Nanotechnology*, ed. H.S. Nalwa, American Scientific Publishers, Vol X, 2003, pp. 1-17
- [22] Saito Y., N. Sasaki, H. Moriya, A. Kagatsume and S. Noro, "Parameter Optimization of Tersoff Interatomic Potentials Using Genetic Algorithms", *JSME International Series A*, Vol. 44, No. 2, 2001 pp. 207-213

# Mobile Motion Gesture Design for Deaf People

Haoyun Xue, Shengfeng Qin  
School of engineering and design  
Brunel University  
London, United Kingdom  
xuehaoyun@gmail.com

**Abstract**—In order to efficiently communicate with non-hearing-impaired (NHI) in particular locations in real-time, deaf people need a more intelligent and easy to use tool beyond their sign language. In this paper, we propose a new communication tool, combining British sign language with gesture-based mobile communication technology. It can quickly translate sign language into text, with one mobile phone, through organized vocabularies in context of different particular locations and integrated touch screen with gesture recognition technology. It can also simplify the process of deaf-to-NHI communication, shows meanings quickly and clearly and gets responses immediately. It bridges the gap in communication between deaf and the NHI. In order to reach this goal, we have designed a set of mobile motion gesture for testing our ideas and further development.

**Keywords**- *British sign language; mobile sign language; deaf people; motion gesture; inclusive design.*

## I. INTRODUCTION

In the UK there are around 9 million people who are deaf or hard of hearing, which means 1 in 7 of the population is deaf. And about 3 in every 1000 children are born deaf [1].

Loss of hearing can cause people to become isolated and lonely, having a tremendous affect on both their social and working life. Society and the government all pay attention to deaf people. For example, specific jobs, social work and “The Royal National Institute for Deaf People” all offer help for deaf people to ease their isolation [2]. In this research, as figure 1 shows, the research focus is on the communication problems of deaf people rather than their social, work and health problems.

There are several tools for deaf people to communicate with. One of them is British Sign Language (BSL). In the UK, there are 50,000-70,000 people who prefer to use British Sign Language to communicate. Sign Language is a visual means of communication using gestures, facial expression and body language, which is used mainly by people who are deaf or have hearing impairments. The BSL was recognized by the UK government as an official minority language in 2003. This has led to an increased awareness of the language which now has a similar status to that of other minority national languages such as Gaelic and Welsh [3]. Hoemann and Courtman-Davies [4, 5] said learning sign language is also important for the NHI to communicate with deaf people.

However, due to unfamiliarity with BSL, the NHI can hardly understand what deaf people mean. From our primary research, we found that nearly 80% of deaf

connections are non verbal. BSL is not a fast and efficient way for deaf people to communicate with the NHI.

We also found that, in general, writing on paper is extensively used within face to face communication between deaf and the NHI. 5 sheets of A4 paper need to be consumed in 20 minutes. The consumption of paper is large for deaf people and unsustainable for the world. Handwriting needs to be considered as a crucial factor to an efficient conversation. Therefore, there is an urgent need to develop an easy communication tool for deaf to have a conversation with the NHI.

On the other hand, nowadays, gesture recognition applications on a mobile phone are ready to be explored. Nokia has launched a mobile phone for the deaf [6]; Nintendo’s Wii allows users to become more engaged in video games with gesture recognition techniques [7]; flip, tilt, tap, snap, and shake mobile phones can easily be recognized by mobile phones and translated to different meanings in different particular situations [8-13].

Mobile phones as the most pervasive wearable computers contain various sensors, such as accelerometers and microphones, as well as actuators in the form of vibro-tactile feedback. To ensure a fast adoption rate of gesture recognition as a ubiquitous input mechanism, technologies already available in mobile phones should be utilized. Features like accelerometer sensing and vibro-tactile feedback are readily available in high-end mobile phones, and this should filter through to most mobile phones in the future.

Nowadays, there is no existing telecommunication system for deaf people to efficiently communicate with the NHI in common locations such as restaurants and coffee shops. Conversation with paper appears to be boring and less interactive. Even with a mobile, current text input method, letters one by one, is quite slow. There is a need to design mobile motion gesture to translate the gestures into letters or words, even sentences, making it easier and quicker to achieve a faster, more natural and interactive conversation.

Our research aims to integrate BSL and gesture based mobile phone interaction into a fast gesture-based mobile communication system for deaf people with one mobile phone. To achieve a fast and clear way of communication, it is designed to directly translate gestures into whole meaningful sentences related to particular common location (context) rather than words or letters. We used interview and observation methods as primary research to obtain and organize particular common location vocabularies, and the literature review as secondary



research to combine British Sign Language with mobile gesture recognition technology. The research framework is shown in Figure 1, clearly demonstrating the related technology, research methods and design process.

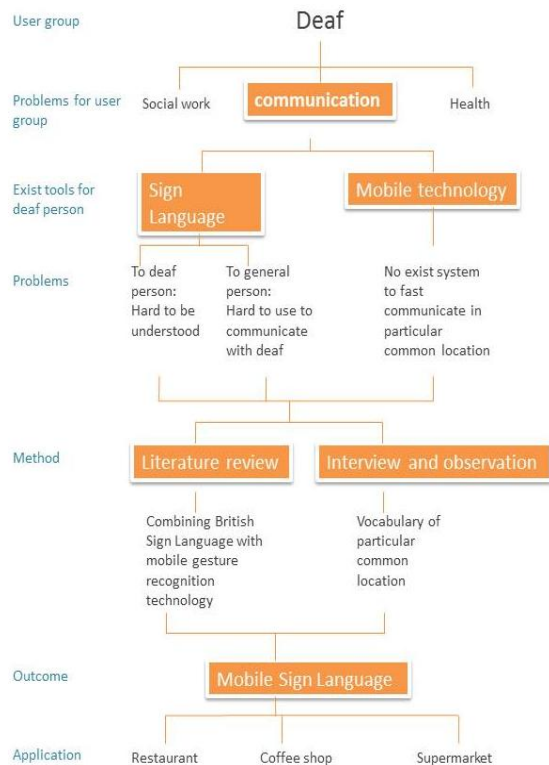


Figure 1 Framework

## II. RELATED WORK

Related research is concerned with gesture-based mobile recognition technology and the design for deaf.

### A. Design for deaf people

Nowadays, design for deaf people based on sign language and mobile phone is developing.

In 1999, Nokia started to design a mobile for deaf people. The Royal National Institute for the Deaf (RNID) has joined forces with Nokia to bring a more portable text phone to the deaf. The partnership was working to adapt the Nokia 9110 Communicator to handle a Talk-Type service. Telephones for the deaf may initially sound like a bit of a non starter, but the technology is actually very simple. Between two text phones, users simply use a combination of plain language text and codes. There is also a text to voice service, sponsored by BT which operates much like a real time paging service. At present, the Nokia 9000i communicator can use the text loop system, on the Orange network. [13]

In 2007, in Britain, the Mobile Phone Sign Language Dictionary was created. This great new piece of innovative sign language technology was developed by the Centre for Deaf Studies at the University of Bristol. Individuals who use sign language are able to download a sign quickly and effectively via their mobiles in 'real time'. One of the great problems for deaf people who use sign language is that until recently there has been no easily

accessible way of gaining a sign language translation of an English word. If therefore, someone is having difficulty in understanding something in English in a particular situation, then the word in hand, can be entered into the dictionary and the sign language equivalent can be accessed. It is extremely easy to access sign language signs using this method. The individual needing the sign should set their mobile phone address page to [www.mobilesign.org](http://www.mobilesign.org). Once this has been done, then an A-Z list of British Signs will come up on the display of the mobile. The individual using the service, will then be able to select the word which they wish to be translated into sign language and the respective sign will be displayed [14].

In 2010, Mobile sign language was being developed at the University of Washington. So far, hearing-challenged consumers have used video chat on PCs. For mobile phones, they have to sent text messages. But that can be limiting because it does not convey emotions, voice inflections or body language. The iPhone 4, HTC Evo and Samsung's Epic 4G phone have front-facing cameras for video conferencing. But video chat on these devices can be too much of a bandwidth hog. Researchers at the University of Washington in Seattle are building the first mobile devices to effectively transmit American Sign Language via compressed video over a 3G cellular network. The aim is to provide real-time video cellular communication for deaf people. The problem they are trying to solve is to optimize the compressed video for sign language so that it can be transmitted on a 3G network, instead of requiring faster 4G network speeds. By increasing image quality around the face and hands, they have brought the data rate down to 30 kilobytes per second. Mobile ASL also uses motion detection to identify whether a person is signing or not so it can help extend the phone's battery life during video use [15].

Recently, the Reach 112 project aims to open up telecommunications to all people. In the UK, deaf users are able to reach 999 services directly through real-time text call (from January 2011) [16].

### B. Mobile gesutre recognition technology

Gesture recognition is an ideal example of multidisciplinary research. There are different tools for gesture recognition, based on the approaches ranging from statistical modeling, computer vision and pattern recognition, image processing, connectionist systems, etc. To mobiles, because they contain various sensors, such as accelerometers and microphones, as well as actuators in the form of vibro-tactile feedback, it seems gesture could be more easily recognized.

In 2005, Samsung launched the gesture controlled cellphone SPH-S4000 and SCH-S400 for Korea. Each of these clamshells is capable of recognizing your convulses for functions such as game play or for skipping or backpedaling MP3 tracks [17].

Recently, Niezen and Hancke [7], based on the mobile accelerometer-based techniques, allow gesture recognition as ubiquitous input for mobile phones. Different gestures such as touching, tapping, shaking, striking and double flipping a phone were explored separately [18]. These

gestures could be understood by the mobile phone. Synthesis of touch and motion gesture techniques also offers great support to the design [19].

### C. Summary

From our research, we can see there is no existing telecommunication system for deaf to be involved in a conversation with the NHI through a faster and more interactive way. Meanwhile, there are no existing applications to directly translate sign language into text to communicate due to the existing mobile gesture recognition techniques and the complexity of sign language still not matching. In our design, we try to fill the gap between existing technology and complex sign language through simplifying the mobile gesture and combining multiple mobile input technologies.

## III. METHOD

In order to build a bridge between existing technology and the complex sign language and to achieve a fast communication with the NHI in particular in common locations, there are two challenges: (1) find out particular vocabularies in different locations, and (2) design a set of gestures which are related to British Sign Language and easy to learn and use. Primary research has been done to solve challenge 1. Literature review has been carried out to solve challenge 2.

### A. Primary research

The aim is to create vocabulary and sentence database through well organized vocabularies. To achieve this aim, questionnaires to deaf and an interview with experts in the deaf center in the University of Bristol have been done to find out the main answers to the questions: which particular locations deaf people always needs to talk to the NHI and generally how they communicate with the NHI. Then, we interviewed staff and customers in these locations to find out the main answers to the questions: what do customers ask most, aiming to create the vocabulary database.

Through our primary research of the deaf, nearly 80% of deaf connections are deaf or sign language users. That means communication limits their connections. 60% of deaf people said they only communicated with the NHI in urgent scenarios in particular in common locations, such as restaurants, coffee shops and supermarkets.

We focus on these locations to explore the vocabularies. Here Figure 2 shows the organized result of scenarios in restaurants, coffee shops and supermarkets.

location	What customers ask most	What deaf people do
Restaurant	Menu	Point at the paper menu
	Pay the bill	Show money/ card
	Know where the toilet is	Fingerspelling the letter 'T'/ write down words
Coffee shop	Menu	Point at the paper menu
	What is the pin number of the toilet	Point at the toilet
supermarket	"Where is"	Fingerspelling the word/ write down words

Figure 2 Organized results

From the result, we can clearly see that in these locations, the questions and vocabularies (see fig 3) are not very much varied. Most of them are interrogative sentences. For deaf people, if they just need to do a simple gesture and could quickly input the key word, then the conversation would be fast.

Location	Basic vocabularies	Basic sentences
Restaurant	bill, menu, toilet, table, tap water, sauce	Interrogative sentences
Coffee shop	toilet, table	
Supermarket	name of items	

Figure 3 Necessary vocabularies

### B. Literature review

The aim is to design a set of gestures which are related to British Sign Language and easy to learn and use. We need to understand more about British Sign Language and the existing mobile gesture recognition technology through a literature review, to find out the basic gestures in daily life and what gestures can be translated into mobile motion gesture and could be easily understood.



The most difficult part for the NHI is the understanding of fingerspelling, which is a method of spelling a word using hand movements. In our design, therefore, we decided to involve touch screen input method instead of gestures to input the key words.

## IV. DESIGN

The goal of the design is to create a faster and easier way for deaf people to communicate with the NHI. Conversation needs to be quick and could add feeling and expression, that is why this mobile sign language should be fast and easy. It has two design parts. One is to design basic gestures, the other is to organize these basic gestures into conversation.

### A. Design of mobile motion gestures

During this stage, we combine British Sign Language with mobile motion gesture technology to create the basic mobile sign language words. These words are all the most basic and most used words of daily life. Mobile motion could be surely recognized by mobile gesture recognition technology. Figure 4 shows the Mobile Sign Language together with the British Sign Language, which is easy and natural for deaf people to remember and use.

British Sign Language	Mobile Sign Language	Description
		You: snap the mobile once to outside











		Me: snap the mobile once to inside
		What/ when/ where/ who/ how/ why: circle the mobile
		Yes: shake the mobile twice like nodding the head
		No: snap the mobile twice like shaking the head
		Here/there: Turn down the mobile and downwards in front of body

Figure 4. Mobile Sign Language

### B. Organize gesture into conversation

We organized all this information and designed this process of conversation for deaf people to shorten the conversation time and to be clearly understood. Showing below is the contrast of before and after using the system in a particular conversation.

Persona:


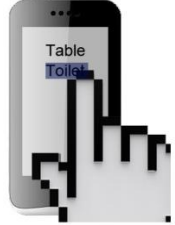


Name	Ben
Age	17
Problem	Hearing impairment
Location	Coffee shop
Situation	To ask what is the pin number for the toilet

Conversation process before:

Ben	Staff
Hi (wave hand to ask staff to pay attention)	Can I help?

What is the pin number of the toilet? (fingerspelling the toilet, hands move fast and seems full of emotion)	What??
Try to find a pen and paper to write down	Wait and confuse
Write down "Pin number of toilet?"	Oh, hand gesture to tell the number, one by one, slowly. At last, go to toilet together with Ben, to help him input the pin number

Conversation process after:

1 Input the word "Toilet"		Screen shows letters to select. Select "T"
		Screen shows the related words. Select "Toilet"
2 input the interrogative sentence		Circle the mobile to start the interrogative sentence
3 the screen shows the related information to the "Toilet"		Choose the right sentence on the screen

Ben	Staff
Hi (wave hand to ask staff to pay attention)	Can I help?
Input the whole sentence through 3 steps	Oh, it's 932814, key in the number with that mobile
Thank you (smile)	Happy ending

The real-time conversation is successfully done through just 3 steps. The NHI could clearly know what deaf people want rather than guessing.



This input method goes through 1) input first letter of the key word by touch screen; 2) screen shows the limited words related to this location; 3) input a typical sentence by motion gesture.

## V. CONCLUSION AND FUTURE WORK

This design provides a successful and fast way for deaf to communicate in particular in common locations with the NHI. It connects existing technology and human needs (Sign Language translation). It can also be used by the NHI to input text with a new method (mobile motion input) and for them to understand basic sign language. It includes all people who understand English rather than sign language which is different from country to country.

In the future, further development works will be carried out. It will combine touch screen technology and mobile accelerometer-based techniques. Lastly, gesture based interface will be programmed.

When the system is created, tests will be done among deaf people. The aim is to test whether it is easy for them to learn and use and to optimize the system.

In the end, collaboration with existing designs for deaf people (such as the mobile phone sign language dictionary, mobile ASL) will be explored.

## REFERENCE

- [1] Website: Fdp.org.uk for deaf person  
<http://www.fdp.org.uk/>
- [2] M. Oliver, and B. sapey, "Social work with disabled people", 3<sup>rd</sup> ed., Palgrave Macmillian, 2006, pp.22-36.
- [3] Website: British Sign Language Resource  
[http://www.british-sign.co.uk/what\\_is\\_british\\_sign\\_language.php](http://www.british-sign.co.uk/what_is_british_sign_language.php)
- [4] M. Courtman-Davies FCST, "Your Deaf Child's Speech and Language" Mary Courtman-Davies, 1979, pp.59-90.
- [5] H. W. Hoemann, "Communicating with deaf people", 2<sup>nd</sup> ed., University Park Press, 1978, pp.xix-xxv, pp. 35-39.
- [6] Website: BBC News Channel, 23 December 2005, "Real-time textng for deaf people"  
<http://news.bbc.co.uk/1/hi/technology/4546924.stm>
- [7] G. Niezen and G.P. Hancke, "gesture recognition as ubiquitous input for mobile phone", International Workshop on Devices that Alter Perception, 2008.
- [8] J. Ruiz and Y. Li, "DoubleFlip: A Motion Gesture Delimiter for Mobile Interaction", UIST'10, New York, New York USA, October 3-6, 2010.
- [9] S. Kallio, J. Kela, J. Manttyarvi, and J. Plomp, "Visualization of hand gestures for pervasive computing environments", Proceedings of the working conference on Advanced visual interfaces, 2006.
- [10] S. Mitra and T. Acharya "Gesture Recognition: a survey " IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, No. 3, pp. 311-324, May 2007.
- [11] C. C Baker, "Explorations in Collaborative Social Video Expression and abstract Mobile communication" Future Internet, 2011.
- [12] S.J. Cho, R. Murray-Smith, C.K. Choi, Y.H. Sung, K.H. Lee, Y.B. Kim, "Dynamics of Tilt-based Browsng on Mobile Devices" CHI EA, 2007.
- [13] Website: Lusy Sherriff, "Nokia launches mobile phone for the deaf Portable Talk-Type system developed with help of RNID", September 2 1999.  
[http://www.theregister.co.uk/1999/09/02/nokia\\_launches\\_mobile\\_phone/](http://www.theregister.co.uk/1999/09/02/nokia_launches_mobile_phone/)
- [14] Website: British Sign Language, deaf news network, May 19, 2007.  
<http://deafnn.wordpress.com/2007/03/19/british-sign-language-mobile-dictionary/>
- [15] N. Cherniavsky, J. Chon, J. O. Wobbroc, R. E. Ladner, and E. A. Riskin, "Activity analysis enabling real-time video communication on mobile phones for deaf users", Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '09), Victoria, British Columbia, October 2009.
- [16] Website: Reach112: responding to all citizens needing help-UK pilot project  
<http://www.reach112.co.uk/>
- [17] S. Bhandari, and Y.K. Lim, "Exploring gestural mode of interaction with mobile phones", CHI 2008, April 5 - April 10, Florence, Italy, 2008.
- [18] A. Scoditti, R. Blanch, and J. Coutaz, "A novel taxonomy for gestural interacion techniques based on accelerometers", Proceedings of the 16<sup>th</sup> international conference on intellingent user interfaces, 2001.
- [19] K. Hinckley and H. Song, "Sensor synaesthesia: touch in motion and motion in touch", Proceedings of the 2011 annual conference on Human factors in computing systems, 2011.

# An Efficient and Secure Authentication Protocol for RFID Systems

Md. Monzur Morshed\* (PhD Student), Anthony Atkins, Hongnian Yu

Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford, ST18 0DF, UK  
m.m.morshed@staffs.ac.uk, a.s.atkins@staffs.ac.uk, h.yu@staffs.ac.uk,

**Abstract--** The use of RFID tags may cause privacy violation of users carrying an RFID tag. Due to the unique identification number of the RFID tag, the possible privacy threats are information leakage of a tag, traceability of the consumer, denial of service attack, replay attack and impersonation of a tag. There are some challenges in providing privacy and security in the RFID tag due to the extremely limited computation, storage and communication ability of passive RFID tags. Many research works have already been conducted using hash functions and random numbers. As the same random number can recur many times the adversary can use the response derived from the same random number for replay attack and it can cause a break in location privacy. This paper proposes an RFID authentication protocol using a static identifier, a monotonically increasing timestamp, a tag side random number and a hash function to protect the RFID system from adversary attacks. The proposed protocol also indicates that it requires less storage and computation than previous existing RFID authentication protocols but offers a larger range of security protection. A simulation experiment is also conducted to verify some of the privacy and security properties of the proposed protocol.

**Keywords-** RFID; security; privacy; timestamp; authentication protocol.

## I. INTRODUCTION

Radio Frequency Identification (RFID) is going to be a part of our everyday life in near future. RFID tags are used in many applications such as in supply-chain management, automation of automobiles, animal tracking, healthcare industry, highway toll collection etc [1]. Many large organizations like Wal-Mart, Procter and Gamble, and the United States Department of Defence are deploying RFID systems for proper control and management of their supply chains [2]. Due to the dropping cost and the improvement in the standardization of RFID tag it is emerging as the successor of optical barcode in many places. RFID has some advantages over optical barcode that make it more suitable in automation. A barcode indicates the type of the object on which it is printed but the RFID tag gives a unique serial number that distinguishes the object uniquely from many millions of similar types of products. Another advantage of RFID tag is that it does not requires line-of-sight contact with the readers as in optical barcode.

RFID is a technology to identify objects or people automatically. An RFID tag is standardized as Electronic Product Code (EPC) tag by the organization EPCglobal Inc [3]. An RFID system consists of three components: tag, reader and back-end database [2, 4]. A typical RFID system is shown in Fig.1. An RFID tag is a small and extremely low-priced device consisting of a microchip with limited functionality and

data storage and antenna for wireless communication with the readers. It transmits data in the air in response to the interrogation by an RFID reader. RFID tags can be passive or active depending on the powering technique [5]. In general passive tags are inexpensive. They have no on-board power; they get power from the signal of the interrogating reader [2]. Active tags contain batteries to power their transmission. Active tags can initiate communications and have read ranges of 100 meters or more. Active tags are expensive and physically larger and hence not suitable for many applications. RFID readers are devices used to read or write data from or to RFID tags. Back-end database has information about the tags.

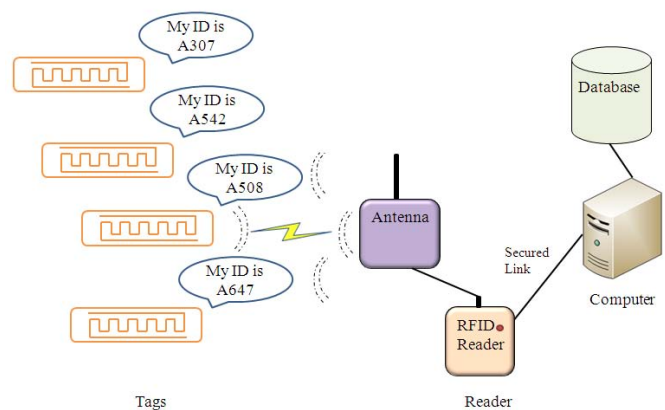


Fig. 1 A typical RFID System

Each RFID tag contains a unique identifier to serve as object identity so that this identity can be used as a link to relate information about the corresponding object. Due to this unique serial number in an RFID tag it is possible to track the tag uniquely and the information in it is vulnerable to an adversary. Products labelled with RFID tags contain unique identifiers. It allows tracking of persons through the tags they carry without their knowledge [1]. Moreover, implementation of conventional cryptography is not possible in a passive RFID tag due to its limited processing capability and memory limitations [6]. The major privacy and security issues in a RFID system are as follows:

- **Information leakage:** In a typical RFID system, a tag has a unique identifier that is transmitted to the reader. So it can be easily identified with this unique serial number. Due to this unique serial number the information in it is vulnerable to an adversary. For the protection from information leakage, an RFID system needs to provide privacy control so that unauthorized readers cannot access the tags.

- **Location privacy:** If any value can be related or linked to a tag then it is possible to track the tag. If a tag transmits any fixed response to a reader, an adversary may try to distinguish it from other response. If an adversary can do it he can find the location of the user.
- **Impersonation and Replay Attack:** The communication between the tag and the reader is insecure. If an adversary can collect the information during communication from the tag and the reader they can impersonate the tag to explore more information. An adversary can use this information and perform replay attack in the future.
- **Message Interception or Denial of Service (DoS):** An adversary may try to block or prevent authentication between a valid reader and tags. If the adversary can successfully block the transmission he can cause the server and the tag to lose synchronization. The RFID system should be able to handle this to keep the synchronization of the tag and the reader.
- **Backward and Forward Traceability:** If the internal state of the tag is known then it can help to identify the tag interactions of past and future communications.

Many researchers have proposed efficient protocols for RFID systems authentication. These protocols can be classified into two categories. First is the hash function based security protocols [4, 7-15]. Second is the lightweight XOR based security protocols [16-24]. Hash function based protocols mostly use random numbers to make the response anonymous. To make the response more anonymous and reliable two random numbers are used in many protocols. One is from the reader side and the other is from the tag side. In most of the XOR based protocols it requires many rounds to authenticate each other.

Some researchers have proposed privacy and security protocols for RFID systems using varying identifiers [4, 7, 16]. These are secured from most of the attacks. Due to varying identifiers they include the recovery from desynchronization due to incomplete authentication process. However, due to the hash function of the constant identifier it gives a fixed value. If one authentication process is unsuccessful, an adversary can use the response later in the subsequent phases to break the security. In this case the adversary can use the response for impersonation and replay attack and also can break the location privacy. Many hash based protocols used static identifier and secret to ensure privacy and security [8-10].

This paper proposes an Efficient and Secure Authentication Protocol (ESAP) based on the challenge-response method using a one-way hash function, a static identifier, a random number and a timestamp in the RFID systems. The objective of this protocol is to overcome the privacy and security problems of the existing protocols with less storage and computations. In this protocol a monotonically increasing timestamp ensures anonymity of the response. The purpose of the hash function is to give a one-way hash result so that an adversary cannot figure out the input from the output. The purpose of the random number is also to make the response anonymous. The monotonically increasing timestamp ensures unique combination of the hash input that makes the function output more anonymous. This protocol protects the privacy and security of RFID systems of the issues outlined above.

The paper is organized as follows. Section II reviews the related works. In Section III the new protocol is presented. In Section IV the privacy, security and efficiency of the protocol are evaluated. In Section V the simulation results and evaluations are presented. Finally in section VI conclusion is given.

## II. RELATED WORKS

There are varieties of RFID authentication protocols for the privacy and security of the RFID systems. Some protocols work with a static identifier and some works with a varying identifier. The protocols work with static identifiers to protect privacy and security so that they can work better in ubiquitous computing environment [8, 9, 10]. Molnar et al. [8] proposed a private authentication scheme for a library RFID system. It uses a pseudorandom number and a shared secret key by the tag and the reader for efficient authentication. This scheme does not ensure forward security since the tag's identifier and the secret key is static and the random number forwarded is in plain text which can be captured by an adversary.

Rhee et al. [9] proposed a challenge response based on the RFID authentication protocol (CRAP) which is designed for use in ubiquitous computing. However, this scheme requires  $(N/2+1)$  hash functions computations which is impractical for a large number of tags in ubiquitous computing.

Choi et al. [10] proposed a one-way hash based low-cost authentication protocol (OHLCAP), which is suitable for ubiquitous environment. This protocol is vulnerable to impersonation attack and traceability attack due to its counter parameter [15].

Tsudik described an RFID identification protocol that provides a basic level of tag identification using time-stamps [25]. Tsudik proposed two further schemes to provide tag authentication [26]. The schemes use monotonically increasing time-stamps for tracking-resistant tag authentication, and employ a keyed hash function  $f$ .

To protect the RFID tags and the reader in an efficient and effective way varying identifiers are used in many authentication protocols [4, 7, 11-14, 16, 17]. This paper focuses on two protocols using varying identifiers and secret numbers for the authentication and is outlined as follows:

Henrici et al. [7] proposed a scheme which is called the hash-based identifier variation scheme (HIDV). The notations used in this protocol are as follows:

- “DB-ID” Database-identifier
- “ID” Current ID
- “HID” Hash of ID acting as a primary index of the table
- “TID” Transaction number
- “LST” Last successful transaction number
- $\Delta TID = TID - LST$
- “AE” Associated DB entry
- “DATA” A reference to tag data / user data

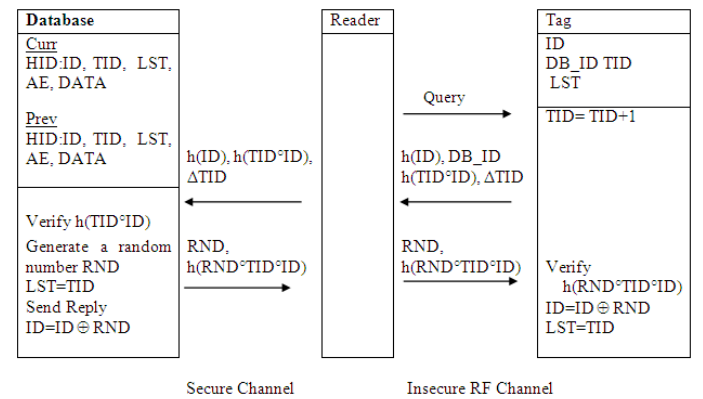


Fig. 2 The HIDV Message Exchange [6]

The operations of the HIDV protocol are shown in Fig. 2. It uses a one-way hash function  $h$  to protect location privacy by changing the  $ID$  after each session. However, if any authentication session is unsuccessful it replies with the same hashed  $ID$  again for which it opens up the vulnerability for

tracking and location privacy [11]. In addition, the proposed scheme is not secure against impersonation attack [10].

Lee et al. [4] proposed a low-cost authentication protocol (LCAP) which simplifies and enhances the HIDV scheme in both efficiency and security. The notations and symbols used in LCAP operation are as follows:

$h : \{0, 1\}^* \rightarrow \{0, 1\}^l$  is a one-way hash function.  $h_R$  is the right half and  $h_L$  is the left half of  $h$ . A hash function is infeasible to invert and the output hides the information of the input [7, 16].  $ID$ :  $ID$  denotes identity of a tag and is a random value in  $\{0, 1\}^l$ .

$r$ : A random number in  $\{0, 1\}^l$ .

Data fields of a tag and a reader are initialized to the following values:

**Tag.** The data field of a tag is initialized to its own  $ID$ .

**Reader:** A reader picks uniformly a random number  $r$ .

The data fields of a back-end database are initialized to  $HaID$ ,  $ID$ ,  $TD$  and  $DATA$ .

**HaID:**  $HaID$  value is the hash value of  $ID$  used for identifying or addressing the tag.

**TD:**  $TD$ -entry is used to trace previous data information of a tag when loss of message occurs in the current session.

**DATA:**  $DATA$  stores the information about an accessible tag.

The back-end database maintains two rows; Prev for the previous session and Curr for the current session. Each row contains  $HaID$ ,  $ID$ ,  $TD$ , and  $DATA$  fields.

The operations of the LCAP protocol are shown in Fig. 3.

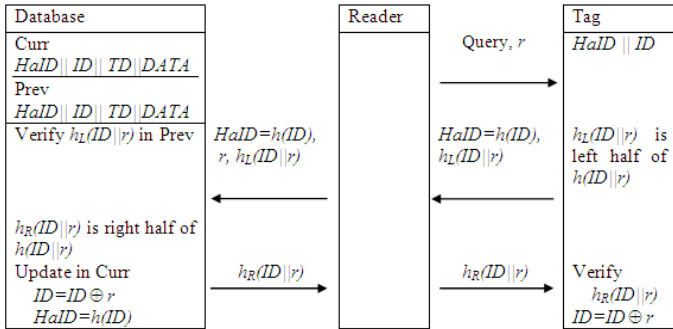


Fig. 3 The LCAP Protocol [4]

It also faces the same problem as in HIDV that is a tag always replies with the same hashed  $ID$  before the next successful authentication which allows tag tracking and breaks the location privacy of the tag [11].

It is an important research consideration to develop a privacy and security protocol for the RFID system that addresses these privacy and security issues and overcomes these problems with the limited storage and computational capacity of an RFID tag. The next section presents the proposed Efficient and Secure Authentication Protocol (ESAP) to overcome the present privacy and security problems.

### III. THE PROPOSED EFFICIENT AND SECURE AUTHENTICATION PROTOCOL

In this section, a new protocol (ESAP) is proposed. This is based on the challenge-response method using a static identifier and a one-way randomized hash function for the RFID systems. This protocol uses a monotonically increasing timestamp to make the response more unidentifiable and anonymous. The notations used in this protocol are as follows:

#### A. Notations

- $h$  A one-way hash function,  $h : \{0, 1\}^* \rightarrow \{0, 1\}^l$
- $l$  The length of an identifier
- $r_1$  Random number in  $\{0, 1\}^l$
- $ID$  Tag identifier
- $X$  Secret number
- $IDX$   $ID \oplus X$ ; it is the search index of the records
- $T_r$  Time stamp generated by the reader
- $T_t$  Last timestamp stored in a tag
- $f_t$  Tag response
- $f_r$  Reader response
- $\oplus$  XOR operator
- $\parallel$  Concatenation operator
- $\leftarrow$  Assignment operator

#### B. System Set-up

**Tag:** Each tag contains the following fields:

$ID$ : Tag identifier

$X$ : Secret number

$T_t$ : Last timestamp

**Reader:** Reader does not contain any fields.

**Back-end Database:** Back-end database contains the following fields:

$IDX$ :  $ID \oplus X$ ; Search index

$ID$ : Tag identifier

#### C. ESAP Operations

When a tag enters into the range of the reader, this can initiate the authentication protocol. The protocol is shown in Fig. 4. The steps in the authentication protocol are as follows.

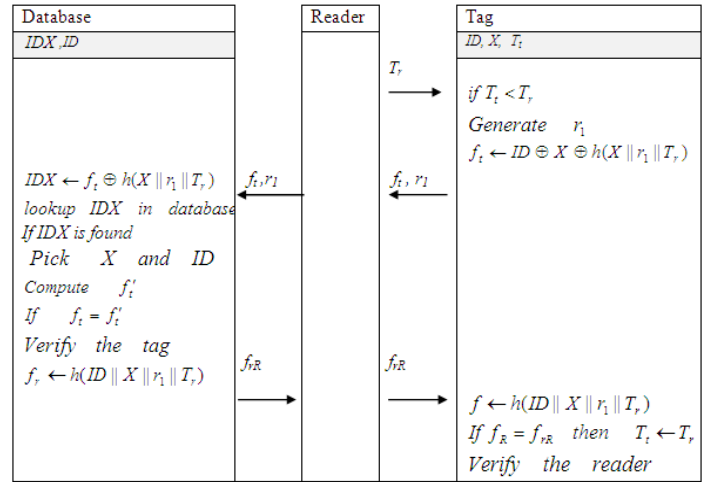


Fig. 4 The Proposed ESAP protocol

1. **Reader:** The reader generates a time stamp  $T_r$  and sends the timestamp to the tag.

3. **Tag:** if  $T_t < T_r$  then

The tag generates a random number  $r_1$

The tag computes  $f_t \leftarrow ID \oplus X \oplus h(X || r_1 || T_r)$

It sends the value of  $r_1$  and  $f_t$  to the reader. The reader then sends  $r_1$  and  $f_t$  to the back-end database.

4. **Database:** The back-end database then computes  $h(X || r_1 || T_r)$  and then it finds  $IDX \leftarrow f_t \oplus h(X || r_1 || T_r)$ .

Lookup  $IDX, ID$  in database and finds  $X \leftarrow IDX \oplus ID$ .

Computes  $f'_t \leftarrow ID \oplus X \oplus h(X || r_1 || T_r)$

if  $f_i = f'_i$  match

Then *authenticates the tag*

Computes  $f_r \leftarrow h(ID \parallel X \parallel r_1 \parallel T_r)$

Finally the back-end database sends

$f_{rR}$  to the reader.  $f_{rR}$  is the right half part of the  $f_r$ .

5. **Reader:** The reader forwards  $f_{rR}$  to the tag.

6. **Tag:** The tag also computes  $f_r$  and checks  $f_{rR}$ . If it matches it authenticates the reader and updates  $T_i \leftarrow T_r$ .

Next we discuss how the protocol works. In this protocol the reader starts authentication by generating a new timestamp  $T_r$  and sends it to the tag. If the timestamp  $T_i < T_r$  then the tag generates a random number  $r_1$  to make the authentication process reliable. The tag then computes the response  $f_i \leftarrow ID \oplus X \oplus h(X \parallel r_1 \parallel T_r)$  and sends  $f_i$  and  $r_1$  to the reader. The reader sends the response and the random number  $r_1$  to the database. The reader at first computes  $h(X \parallel r_1 \parallel T_r)$  and then it computes  $IDX \leftarrow f_i \oplus h(X \parallel r_1 \parallel T_r)$ . If  $IDX$  is found in the database it uses the  $ID$  to calculate the secret  $X$ . The database then calculate  $f_i$  with this values. If it matches with the  $f_i$  received from the tag then authenticate the tag. The database then computes  $f_r \leftarrow h(ID \parallel X \parallel r_1 \parallel T_r)$  and sends the right half  $f_{rR}$  to the reader. The protocol uses one monotonically increasing timestamp to keep the response unidentifiable or anonymous. The tag then computes the  $f \leftarrow h(ID \parallel X \parallel r_1 \parallel T_r)$ . If the right half of this value matches with the received one then the reader is authenticated. The proposed protocol uses a random number for the tag side and a timestamp from the reader side. It makes the response more unpredictable. Moreover the monotonically increasing timestamp also makes the input combination unique and intractable.

#### IV. ANALYSIS OF THE PROPOSED PROTOCOL

To evaluate the proposed protocol privacy, security and efficiency will be analysed.

##### A. Privacy and Security Analysis

The privacy and security of the proposed protocol are analysed against the threats discussed in Section I.

- **Information Leakage:** To be able to obtain any sensitive information from a tag a protocol must be authenticated. In this protocol, to authenticate the system an adversary must know  $ID$  and the hash function to receive any information from the tags. The combination of  $r_1$ ,  $T_r$  and  $ID$  makes the responses so unpredictable that the adversary can only guess the value or use a brute-force technique with an advantage of only  $(1/2^l)$ , which is negligible for data length of 96 bits or more.
- **Location Privacy:** The value of  $f_r$  and  $f_i$  cannot be linked with any particular tag. The protocol ensures location privacy by using new values of  $r_1$ ,  $T_r$  each time. Even if a malicious reader sends the same timestamp  $T_r$  all the times, a tag transmits the refreshed value using  $r_1$ ,  $X$  and  $ID$ .
- **Impersonation and Replay Attack:** When a tag reaches within the range of a reader, the reader queries with a random value to the tag. An adversary may also make a request to a tag with a timestamp. However, without

knowing the  $ID$ ,  $X$ , the hash function an adversary is unable to impersonate. For each session the tag generates a new value of  $f_i$  which is totally indistinguishable and different from other session and subsequently the impersonation and replay attacks are not possible.

- **Message Interception or DoS attack:** It is not possible to detect all the types of DoS attacks. The objective of the protocol is to take action against the vulnerability of a DoS attack and the system should not be desynchronized. The proposed protocol uses a static identifier for the authentication process. If the adversary is able to prevent the last transmission to the tag from the reader then the tag will not authenticate the reader in that session. In the next authentication phase it will use a new random number to authenticate and the reader will send a new timestamp and the process will be continued.
- **Traceability:** An adversary is unable to identify the tag from its response because each time it gives a different value which is non traceable from other responses. This scheme is fully protected from the future forward and backward traceability. The adversary has no control over  $r_1$ , and the combination of  $r_1$ ,  $T_r$  and hash function and also does not know the  $ID$  and secret  $X$ . Consequently, the previous, present and future interactions are all indistinguishable.

##### B. Efficiency Analysis

Storage, communication and computation cost were considered for efficiency analysis. Two existing authentication protocols OHLCAP [10] and YA\_TRAP\* [26] were compared with the proposed ESAP authentication protocol. These protocols were selected for efficiency comparison since all of them work in ubiquitous environment. OHLCAP and YA\_TRAP\* require a larger storage and computations than ESAP protocol. OHLCAP is also vulnerable to impersonation attack. The ESAP protocol shows improved performance as shown in Table I because it requires less tag side and database side storage than other protocols. The storage requirement for the tag and the database are  $3l$  and  $2l$  respectively. The protocol requires less hash function in both tag and database. YA\_TRAP\* cannot give protections from some of the attacks and it requires  $(N/2+1)$  complex functions operations which is costly because the value of  $N$  may be very high and it requires many function computations that will make the protocol slower [26]. Table I gives an overall comparison of the different protocols compared to the proposed ESAP. Another advantage of the proposed protocol is that it requires less data to be communicated from the reader to the tag.

TABLE I EFFICIENCY ANALYSIS

Efficiency Criteria		OHLCAP	YA_TRAP*	ESAP
Storage	Tag	$5l$	$4l$	$3l$
	Reader	-	-	-
	Database	$4l$	$5l$	$2l$
Computation	Tag	$1h+A$	$2h$	$2h$
	Reader	-	-	-
	Database	$1h+e$	$(N/2+1)h$	$2h$
Communication	Tag-to-Reader	$2.5l$	$3l$	$2l$
	Reader-to-tag	$0.5l$	$3l$	$0.5l$

A, e: Operations in a tag and a database respectively except for hash operation



## V. EXPERIMENT RESULTS AND EVALUATION

To validate the proposed protocol ESAP, simulation work has been conducted. The privacy and security protections are ensured with the hash functions, timestamp and random number. A hash function is a one-way function for which information leakage is not possible from the hash response. The simulation is to further verify the protection for impersonation attack, replay attack and location privacy. It is assumed that the adversary will capture a response from the tag or the reader and then subsequently use this response  $10^{11}$  times to impersonate the tag or the reader. It checks the responses  $f_i$  and  $f_r$  if any of them recur more than once for one tag during the attacks by an adversary. If the same response is generated it can be used by the adversary for impersonation and replay attack and the location privacy of the tag may be broken. A simulation program in Turbo C++ compiler is developed. It runs in a desktop computer of Intel (R) Core 2 Duo. Processor speed is 2.93GHz and memory 3.46 GB. The operating System was Windows XP professional. The objective of the simulation program was to check the response for one tag if the response is anonymous. The output of a hash function is the same for the same random number and timestamp. Our objective is to practically ensure unique response for different inputs of random number and timestamp so that attacker cannot use any response it collected and attack later to access the tag or the reader. The program checks to match a response with subsequent responses for a set of random number and time stamp. The number of times the same response generated for the tag response  $f_i$  and the reader response  $f_r$  is given in the Table II. It represents the success of the adversary for  $10^{11}$  attempts for different sizes of secret numbers and data. The experiment was conducted for 16 bits, 32 bits, 64 bits and 96 bits secret and data length. In this experiment there was no match of the response for 64 bits and 96 bits. For 16 bits and 32 bits there were some recurrences of the same response. The reason is that it produced the same response for some other combination of random number and the timestamp. The recurrence of the response for 16, 32, 64 and 96 bits are shown in the table for  $10^{11}$  attempts.

TABLE II ATTACKER'S SUCCESS TABLE

Exp No	Number of Queries to the Tag	Attacker's Success for different data length		
		Data length	Number of Matches	
			$f_i$	$f_r$
1	$10^{11}$	16	1538360	1538360
2	$10^{11}$	16	1550799	1550799
3	$10^{11}$	16	1527728	1527728
4	$10^{11}$	32	20	20
5	$10^{11}$	32	15	15
6	$10^{11}$	32	0	0
7	$10^{11}$	32	0	0
8	$10^{11}$	32	25	25
9	$10^{11}$	32	23	23
10	$10^{11}$	64	0	0
11	$10^{11}$	64	0	0
12	$10^{11}$	64	0	0
13	$10^{11}$	64	0	0
14	$10^{11}$	64	0	0
15	$10^{11}$	96	0	0
16	$10^{11}$	96	0	0
17	$10^{11}$	96	0	0
18	$10^{11}$	96	0	0
19	$10^{11}$	96	0	0
20	$10^{11}$	96	0	0

This experiment shows that during the attempt with 64 and 96 bits data and secret the tag and the reader produced unique response for a tag  $ID$  and the adversary cannot break the privacy and security of the RFID systems.

In this experiment the attacker only tries to track the response in passive mode. It cannot use the previous timestamp and the response to attack the tag, since the tag always checks if the new timestamp is larger than its stored one. The tag does not modify its timestamp until an authentication process is successful. This experiment showed that the protocol is secure for at least 64 bits data and secrets in  $10^{11}$  attempts. The following Table III shows the evaluation summary.

TABLE III ATTACKER'S SUCCESS SUMMARY

Number of Queries	Attacker's Success					
	Data length (16 bits)		Data length (32 bits)		Data length (64/96 bits)	
	$f_i$	$f_r$	$f_i$	$f_r$	$f_i$	$f_r$
$10^{11}$	>0	>0	>=0	>=0	0	0

$f_i$ : Tag Response,  $f_r$ : Reader Response

In this authentication system it is not possible to perform an active attack by the adversary to the tag by using the same timestamp. The reason is that the tag always stores the last timestamp and it does not allow any authentication process until it receives a timestamp greater than the previous one. Due to this monotonically increasing timestamp impersonation and replay attack is not possible. Another advantage of this protocol is that the adversary cannot be successful with arbitrary big fake timestamp since the tag does not update its timestamp unless a successful authentication is performed. This prevents the protocol from DoS attack.

The summary of the privacy and security properties is given in Table IV. The privacy and security properties of ESAP are compared with four other schemes [4, 10, 11, 15, 26]. The four schemes were chosen because all of these protocols involved tag authentication. HIDV and LCAP involve secret update process and other two protocols OHLCAP and YA\_TRAP\* do not support secret update. ESAP is similar to OHLCAP and YA\_TRAP\* since ESAP does not support secret update and all these protocols support authentication in ubiquitous environment. The table shows that the proposed protocol provided protections from all the identified privacy and security threats.

TABLE IV PRIVACY AND SECURITY COMPARISONS

Property	HIDV	LCAP	OHLCAP	YA_TRAP*	ESAP
Information privacy	Y	Y	Y	Y	Y
Location Privacy	N	N	Y	Y	Y
Impersonation	N	A	N	Y	Y
Replay attack	N	Y	N	Y	Y
Message Interception	Y	Y	Y	N	Y
Backward Traceability	N	Y	N	N	Y
Forward Traceability	N	Y	N	N	Y

Y: Protected A: provided under assumption N: Not Provided

Fig. 5 shows the storage comparison with two other ubiquitous RFID privacy and security protocols. Storage requirement in ESAP is less than other protocols. Storage requirements are presented as  $l$  bits. The HIDV and LCAP protocols are not included in storage comparison since they update their ID after each authentication phase.

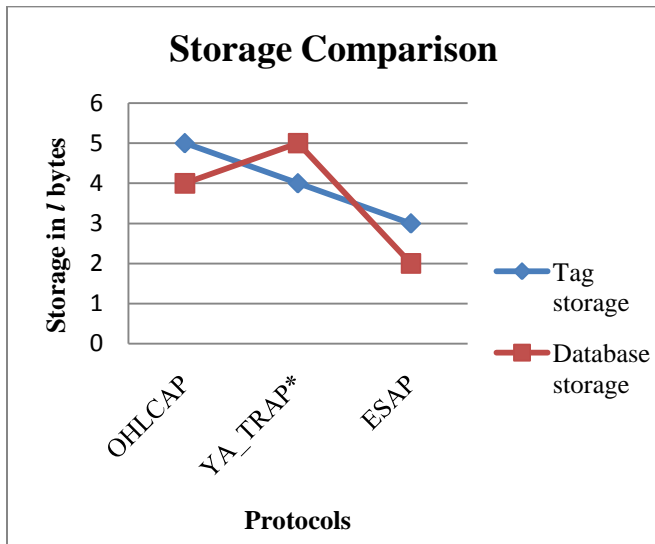


Fig. 5 Storage Comparison

The simulation experiment successfully authenticates the tag and the reader without any privacy and security failure.

## VI. CONCLUSION

A new efficient and secure authentication protocol ESAP has been presented in this paper to protect privacy for low-cost RFID systems. The protocol uses a static identifier to provide effective privacy and security in a ubiquitous environment using hash functions, a timestamp and a random number. The strength of this protocol is the use of a monotonically increasing timestamp and a random number to make the response more unidentifiable. This protocol uses the search index *IDX* to search the tag records in the database. It reduces the tag search time substantially in the database. The simulation experiment also proved that, the responses during the experiment were unique for both the 64 and 96 bits long secret and data length. It is secured from an adversary from all the attacks discussed in Section I. Specific privacy and security protections from an adversary appropriate to simulation experiment were tested and found to be satisfactory. The privacy and security protections were also analyzed and the analysis verified that this protocol is protected from the identified threats. The proposed scheme requires only two one-way hash functions making it highly efficient. The storage requirements for the tag and database are also cost efficient. The comparison outlined in the analysis and experiment result shows that the proposed protocol is secure and efficient in compared to the other protocols. It has practical advantages over these protocols because it is simple and provides a larger range of privacy and security protections. This protocol will be suitable in the RFID systems of healthcare industry, shopping mall and supply chain management etc.

## REFERENCES

- [1] S. Garfinkel, A. Jules and R. Pappu, "RFID privacy: an overview of problems and proposed solutions," *IEEE Security and Privacy*, 3(3): 34-43, May/June 2005.
- [2] A. Jules, "RFID security and privacy: A research survey," *IEEE Journal on Selected Areas in Communication*, 24(2), February 2006.
- [3] EPCglobal Web site, 2005. Referenced 2005 at <http://www.EPCglobalinc.org>.
- [4] S.M. Lee, Y.J. Hwang, D.H. Lee and J.I. Lim, "Efficient Authentication for Low-Cost RFID systems," *ICCSA05*, vol. 3480 LNCS, pp.619-629, Springer-Verlag, 2005.

- [5] R. Want, "An Introduction to RFID Technology," *IEEE Pervasive Computing*, vol. 5, pp. 25 – 33, 2005.
- [6] B. S. Prabhu, X. Su, H. Ramamurthy, C. Chu and R. Gadh, "WinRFID – A Middleware for the enablement of Radio Frequency Identification (RFID) based Applications," *UCLA - Wireless Internet for the Mobile Enterprise Consortium (WINMEC)* 420 Westwood Pl., Los Angeles CA 90095.
- [7] D. Henrici and P. Muller, "Hash-based enhancement of location privacy for radio-frequency identification devices using varying identifiers," In R. Sandhu and R. Thomas, editors, *International Workshop on Pervasive Computing and Communication Security -PerSec 2004*, pages 149–153, Orlando, Florida, USA, March 2004. IEEE Computer Society.
- [8] D. Molnar and D. Wagner, "Privacy and security in library RFID: Issues, practices, and architectures," In B. Pfizmann and P. Liu, editors, *Conference on Computer and Communications Security - ACM CCS*, pages 210–219, Washington, DC, USA, October 2004. ACM Press.
- [9] K. Rhee, J. Kwak, S. Kim and D. Won, "Challenge-Response Based RFID Authentication Protocol for Distributed Database Environment," *SPC 2005*, LNCS 3450, pp. 70-84, 2005.
- [10] E.Y. Choi, S.M. Lee and D.H. Lee, "Efficient RFID Authentication Protocol for Ubiquitous Computing Environment," *Embedded and Ubiquitous Computing*, vol.3832, pp.945-954, 2005.
- [11] H. Chien and C. Chen, "Mutual authentication protocol for RFID conforming to EPC class 1 generation 2 standards," *Computer Standards & Interfaces*, 29(2):254–259, February 2007.
- [12] M. Ohkubo, K. Suzki and S. Kinoshita, "Cryptographic approach to "privacy-friendly" tags," In *RFID Privacy Workshop*, MIT, MA, USA, November 2003. <http://www.rfidprivacy.us/2003/agenda.php>.
- [13] T. Dimitriou, "A lightweight RFID protocol to protect against traceability and cloning attacks," In *Conference on Security and Privacy for Emerging Areas in Communication Networks - SecureComm*, pages 59–66, Athens, Greece, September 2005. IEEE.
- [14] M.E. Hoque, F. Rahman and S.I. Ahmed, "Supporting Recovery, Privacy and Security in RFID Systems Using A Robust Authentication Protocol," *Proceedings of the 2009 ACM symposium on Applied Computing, SAC'09*, Honolulu, Hawaii, USA. pp.1062-1066.
- [15] Ha, J, Moon, S, Nieto, JMG and Boyd, C "Security Analysis and Enhancement of One-Way Hash Based Low- Cost Authentication Protocol", *Emerging Technologies in Knowledge Discovery and Data Mining*, vol.4819, pp.574-583, 2007.
- [16] B. Song and C. J. Mitchell, "RFID authentication protocol for low-cost tags," In *WISEC*, pages 140-147, 2008.
- [17] B. Song, "RFID Tag Ownership Transfer," In *4th Workshop on RFID Security (RFIDsec 08)*, Budapest, Hungary, July 2008.
- [18] N.J. Hopper and M. Blum, *Secure human identification protocols*, *Advances in Cryptology – ASYACRYPT'2001*, Lecture Notes in Computer Science, vol. 2248, Springer, 2001, pp.52–66.
- [19] A. Juels and S. Weis, *Authenticating Pervasive Devices with Human Protocols*, *Proceedings of CRYPTO'05*, Victor Shoup (Ed.), Springer-Verlag, LNCS 3261, pp. 293– 308, 2005.
- [20] H. Gilbert, M. Robshaw and H. Sibert, *An active attack against HB+ – a provably secure lightweight authentication protocol*. Manuscript, July 2005.
- [21] J. Katz and J.S. Shin, *Parallel and concurrent security of the HB and HB+ protocols*, *Cryptology ePrint archive*, Report 2005/461, 2005, <http://eprint.iacr.org>.
- [22] J. Bringer, H. Chabanne and E. Dottax, *HB++: a lightweight authentication protocol secure against some attacks*, in: *IEEE International Conference on Pervasive Services, Workshop on Security, Privacy and Trust in pervasive and Ubiquitous Computing – SecPerU*, 2006.
- [23] S. Piramuthu, *HB and related lightweight authentication protocols for secure RFID tag/reader authentication*, in: *COLLECTeR Europe Conference*, Basel, Switzerland, 9–10 June, 2006.
- [24] J. Munilla and A. Peinado, *HB-MP: A further step in the HB-family of lightweight authentication protocols*, *Computer Networks* 51 (2007), p.2262-2267, 2007.
- [25] G. Tsudik, *YA-TRAP: Yet another trivial RFID authentication protocol*. In *Fourth IEEE Annual Conference on Pervasive Computing and Communications -- PerCom 2006*, pages 640-643, Pisa, Italy, March 2006. IEEE Computer Society.
- [26] G. Tsudik, *A family of dunces: Trivial RFID identification and authentication protocols*. In N. Borisov and P. Golle, editors, *Privacy Enhancing Technologies, 7th International Symposium -- PET 2007*, volume 4776 of *Lecture Notes in Computer Science*, pages 45-61, Ottawa, Canada, June 2007. Springer-Verlag, Berlin.

# Design of Interference Aware ZigBee Building Monitoring Network

Fang Yao and Shuang-Hua Yang\*  
Computer Science Department  
Loughborough University  
Loughborough, UK  
F.Yao & S.H.Yang@lboro.ac.uk

**Abstract**— ZigBee is a new standard developed by the ZigBee Alliance for providing low cost, low power consumption wireless solutions. By integrating with dedicated sensors, ZigBee mesh network can be easily applied for large-scale building monitoring applications. However, the wide adoption of wireless products (e.g. IEEE 802.11 networks) makes the operations of ZigBee network easy to be affected as they all use 2.4 GHz ISM band. Particularly, interference existing inside a building environment includes static interference and dynamic interference, which pose challenges for ZigBee network communications. In this work, we present a set of guidance for deploying ZigBee devices to avoid static interference from Wi-Fi and propose a set of strategies to mitigate dynamic interference from Wi-Fi in a building environment. The mitigation guidance and strategies are derived from our experimental work and take various factors, including position selection, channel allocation, and dynamic interference response, into consideration.

**Keywords**- ZigBee; building monitoring; interference

## I. INTRODUCTION

The rapid development of ZigBee [1] technologies has been leading the use of wireless sensor networks (WSN) in building environment monitoring [2]. Using sensor network to implement building environment monitoring is traditionally achieved by wired systems. However, they raise many issues such as high installation cost, high maintenance cost, and difficult to access, etc. The ZigBee technologies can easily resolve these issues by enabling sensor nodes with the capability of constructing a self-organizing and self-healing wireless network. However, due to the wide popularity of wireless products, it is inevitable for ZigBee wireless devices to co-exist with other wireless systems that work on the same 2.4 GHz ISM band. Consequently the performance of the ZigBee wireless devices will be degraded.

The most serious interference problem in building environment monitoring is the coexistence of ZigBee monitoring network and IEEE 802.11 network, or called Wi-Fi. There are often multiple IEEE 802.11 networks existing inside a large-scale building such as a commercial or government building. For the convenience of employee or visitors accessing networks (e.g. Internet, Intranet) at offices, many IT services have made Wi-Fi networks as a standard accessory of any building network system. For

the security and independency reason, it is quite commonly to have multiple IEEE 802.11 networks to work in a same area as required by different organizations, or departments. When a ZigBee sensor node is located in the close vicinity of the IEEE 802.11 wireless routers or access points, the strong output power of 802.11 signals (typically at 20 dBm) will cause considerable interference on the ZigBee receiver. Related research made in [3] [4] [5] have pointed out that when the physical distance of a ZigBee device from a IEEE 802.11 transmitter is less than 8 meters, or the center frequency employed by the ZigBee network is close to the IEEE 802.11 communication channel (i.e. the frequency offset is less than 7MHz), the packet error rate of ZigBee communications will dramatically increase. These criteria might be not hard to be satisfied in the deployment of a ZigBee sensor network inside a building as most 802.11 routers are static after installation. However, some 802.11 devices with features of high mobility (e.g. 802.11b/g/n network adaptors equipped in laptops) could generate serious dynamic interference. In this work, we have analyzed the relationship between ZigBee systems and Wi-Fi interference at a system level, and proposed a set of mitigation strategies in order to achieve a better design for building environment monitoring. The rest of the paper is organized as follows. Section II overviews the ZigBee technologies and the features of building environment monitoring. The analysis of static and dynamic interference affecting ZigBee network is given in Section III. Section IV illustrates how to deploy ZigBee wireless sensor networks to avoid static interference from Wi-Fi network. Section V presents a new mitigation strategy for ZigBee systems to avoid and mitigate the effect of dynamic interference. The experimental evaluation tests are described and analyzed in Section VI. Section VII concludes the paper.

## II. OVERVIEW OF ZIGBEE AND BUILDING ENVIRONMENT MONITORING

### A. ZigBee Standard

ZigBee is a worldwide standard. The main objective of ZigBee is to provide an open standard suitable for wide range of applications that perform monitoring or control functions. ZigBee standard is an enhancement of IEEE 802.15.4 standard. Figure 1 illustrates the architecture of ZigBee protocol.



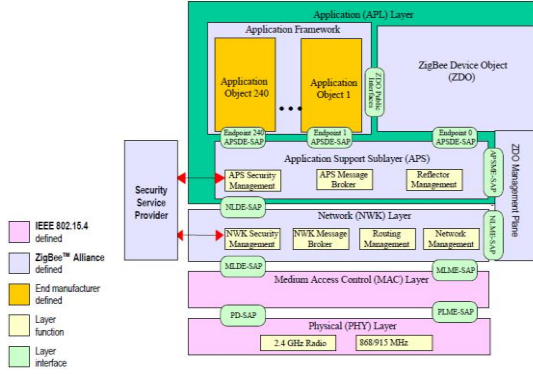


Figure 1. ZigBee stack architecture [1]

As shown in Figure 1, the ZigBee stack consists of four layers: Physical and MAC layers are to provide ZigBee systems with the capability of low power consumption wireless communication. Network layer is responsible for network topology construction and routing protocol implementation. ZigBee standard supports star, tree, and mesh topologies. Mesh topology is the most suitable one for the use in building environment monitoring due to its large-scale deployment. ZigBee standard defines three types of devices: ZigBee coordinator responsible for starting and maintaining network, ZigBee router responsible for relaying network messages, and ZigBee end device responsible for implementing sensing or control tasks. Application layer is mainly for developers to define the content of applications (e.g. reading sensors, controlling actuators).

### B. Building Environment Monitoring

A ZigBee wireless sensor network deployed for monitoring environment conditions inside a building should be able to adopt hundreds of sensor nodes. Figure 2 illustrates a typical ZigBee monitoring network.

A ZigBee coordinator and multiple ZigBee routers compose a ZigBee mesh network deployed inside building in Figure 2. Each router can freely talk to other routers within its radio communication range. ZigBee end devices integrated with a number of environment sensors (e.g. temperature and humidity sensors) join the ZigBee network through the nearest ZigBee routers. ZigBee end devices can be installed at any position where at least one ZigBee router device is in its radio range. When sensing information is required by the monitoring officers the ZigBee end device with the sensing data will send data to its parent router device, then the router passes the sensing data to the destination by using a routing protocol.

## III. INTERFERENCE IN ZIGBEE MESH NETWORK

For a ZigBee mesh network applied in the context of real-life application, the interference analysis should take comprehensive considerations, including static and dynamic interference.

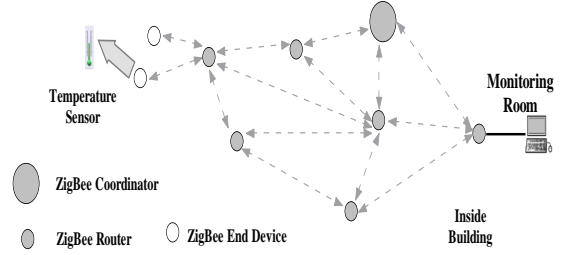


Figure 2. ZigBee wireless sensor network deployment inside building

### A. Static Interference

Figure 3 illustrates a deployment of multiple access points for meeting the requirement of allowing multiple users to access an Ethernet network in a large coverage area. Laptops connect to the Ethernet through wireless access points and usually work under “downlink mode”, which means that most 802.11 communications are issued from the access points to the laptops.

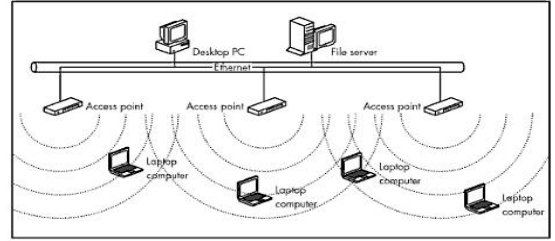


Figure 3. Multiple Wi-Fi network deployment [6]

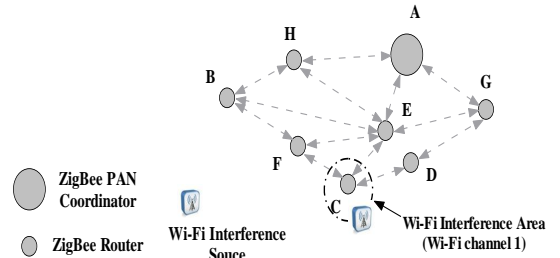


Figure 4. Interference scenario in ZigBee mesh network

A possible static interference scenario is illustrated in Figure 4, where a Wi-Fi interference source is located close to ZigBee device C. Assuming device C being interfered, if device F needs to send data to device G, the possible routes could be F->C->D->G, F->B->E->G, and F->E->G. The first route will be discarded by the routing protocol as device C is not available. However, if more ZigBee routers are affected by Wi-Fi interference source, ZigBee communications in mesh network are probably failed due to the failure of route establishment. It is inevitable to have some ZigBee devices coexisting with Wi-Fi access points.

Wi-Fi transmitter has a much higher output power than ZigBee transmitter. The initial output power of Wi-Fi transmitter is around 100 mW, 100 times of the standard ZigBee transmitter output power (i.e. 1 mW). As the increment of propagation distance, Wi-Fi signal power

will attenuate. If the remaining Wi-Fi power reaching a ZigBee receiver is less than an allowed noise level, the interference effect on a ZigBee network can be ignored.

Channel allocation is another way to avoid the Wi-Fi interference to a ZigBee network. Although a ZigBee network can use up to 16 communication channels defined at the 2.4 GHz band, it can only use one communication channel at a given time and does not hop its frequency. When the distributed Wi-Fi routers utilize non-overlapping channel settings, only a small number of ZigBee channels are unaffected. However, most of the energy of Wi-Fi signal concentrates on the central frequency of the employed Wi-Fi channel, a certain frequency separation between a Wi-Fi channel and a ZigBee channel can effectively reduce interference, which is critical for a ZigBee network to survive when it coexists with multiple Wi-Fi networks.

#### B. Dynamic Interference

Wi-Fi devices were originally designed for meeting the mobility purpose. People can easily remove any installed access points or add new one in a different location when it is necessary. For example, a temporal Wi-Fi access point may be set up in a room as requested by a meeting include participants from external organizations. These unexpected interference devices were not scheduled in the original Wi-Fi deployment and may generate serious interference on any ZigBee system deployed in the area. Due to the wide distribution of ZigBee sensor network, the ZigBee PAN coordinator could be hundreds of meters away from the devices being affected by dynamic interference source, and may not be sensitive to the emergence of dynamic interference area. Furthermore, the duration and level of interference caused by any dynamic interference source is unknown to the sensor network. Consequently the performance of the whole ZigBee network will be affected under the dynamic interference.

#### IV. DEPLOYMENT OF WIRELESS SENSOR NETWORK

The locations of most Wi-Fi routers in a building environment are fixed during day-to-day operation. Therefore ZigBee devices can be optimally deployed to avoid interference caused by Wi-Fi by keeping a safe distance away from those Wi-Fi transmitters, and employing a suitable communication channel. This section will illustrate the way to deploy ZigBee wireless routers by conducting a three-step experiment.

Assume that the pre-installed Wi-Fi routers are static and have a limited interference range. If the distance between a Wi-Fi router and a ZigBee receiver is fixed at a safe value (greater than 8 meters in terms of the experiments made in [3] [4] [5]), the success rate of ZigBee communications will vary only with the distance between a ZigBee transmitter and the ZigBee receiver and the frequency separation between ZigBee and Wi-Fi networks. Figure 5 illustrates the set-up of our experiments where the safe distance is set at 10 meters away from a Wi-Fi interference source.

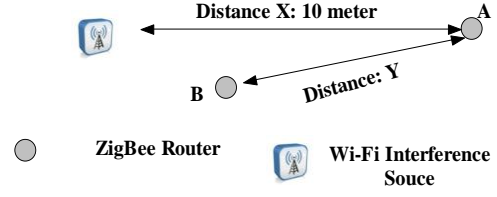


Figure 5. Device deployment in static interference avoidance

In the above experiment set-up, the location of the Wi-Fi transmitter is fixed at 10 meters away from a ZigBee device (device A), and continues to broadcast Wi-Fi signals. ZigBee devices A and B are all ZigBee routers, and act as ZigBee receiver and transmitter respectively. The experiment is to measure the success rate of data transmission on different ZigBee channels between devices A and B when the distance between them, i.e. distance Y, varies. The experiment is divided into three steps: A) measure interfering energy distributed on different ZigBee channels, B) measure ZigBee signal strength falling within a ZigBee receiver bandwidth, and C) measure ZigBee communication success rate under Wi-Fi interference with different distance Y and on different ZigBee Channels.

Experiment A: The Wi-Fi transmitter was set working on channel 11 (2462 MHz). An energy detector compliant with ZigBee standard was placed at the side of device A. The detector was listening on 16 ZigBee channels and recorded the highest energy level for all 16 channels. It is obvious that the detected Wi-Fi interfering energy level is higher at the ZigBee channels (channels 22 and 23) which are closer to the center frequency 2462 MHz of the working Wi-Fi channel (channel 11). Table I shows the results.

TABLE I. INTERFERENCE ENERGY LEVEL CAUSED BY WI-FI SIGNAL ON ALL ZIGBEE CHANNEL

ZigBee Channel (Centre Frequency)	11 (2405 MHz)	12 (2410 MHz)	13 (2415 MHz)	14 (2420 MHz)	15 (2425 MHz)	16 (2430 MHz)	17 (2435 MHz)
Measured Energy (dBm)	-95	-96	-96	-92	-88	-82	-80

18 (2440 MHz)	19 (2445 MHz)	20 (2450 MHz)	21 (2455 MHz)	22 (2460 MHz)	23 (2465 MHz)	24 (2470 MHz)	25 (2475 MHz)	26 (2480 MHz)
-76	-65	-59	-36	-26	-20	-38	-57	-76

Experiment B: This measurement is to detect the ZigBee signal strength falling within a ZigBee receiver's bandwidth when no Wi-Fi interference is present. We choose channel 23 (2465MHz) as the ZigBee device working channel. Distance Y changes from 2 meters to 10 meters. The result is summarized in Table II, which gives an ideal ZigBee signal strength for various distances between the ZigBee transmitter and receiver.

TABLE II. ZIGBEE SIGNAL STRENGTH ON A RECEIVER WITH DIFFERENT DISTANCE FROM THE TRANSMITTER

Distance Y (meter)	2	4	6	8	10
Measured Energy Level (dBm)	-47	-51	-53	-55	-57

Experiment C: This measurement is designed to determine the success communication rate under a fixed Wi-Fi interference with various distances between a ZigBee transmitter and a receiver when they are working on various ZigBee channels.

A constant Wi-Fi traffic is generated by a software packet generator and broadcast around. ZigBee device B was set to send data packet to device A with a fixed packet rate at 200 packet/second. The ZigBee packets contained a fixed payload length at 50 bytes. This ZigBee transmission lasts for 50 seconds, and 10,000 packets in total have been sent out. The received packet number is measured in device A. The test results are shown in Table III. When ZigBee network works on channels 23 and 24, the success rates are significantly poor even distance Y was set at 2 meters. When ZigBee network works on channel 25 and distance Y was set at 8 meters, the success rates achieve 89.9% although the ZigBee signal strength is 2 dB higher than Wi-Fi energy. Once distance Y decreases to 6 meters and the ZigBee signal strength is 4 dB higher than Wi-Fi energy, the success rates become 99.67%. It accords with the simulation result conducted by the IEEE 802.15.4 standard that said when the ZigBee signal strength measured on the receiver is about 4-5 dB greater than the noise level, the packet error rate will be less than 1%. This experiment demonstrates that the frequency separation is more important than signal strength in the avoidance of interference.

The conclusions drawn from the above experiments A, B, C are that if a ZigBee router can be installed 10 meters away from the nearby Wi-Fi router, and the ZigBee network employs the communication channel whose center frequency is at least 13 MHz away from the Wi-Fi communication channels, other ZigBee devices can be safely installed around the ZigBee router within a range of 8 meters.

## V. DYNAMIC INTERFERENCE DETECTION AND MITIGATION

Unexpected interference sources may operate on any ZigBee channel. If multiple interference sources employing different Wi-Fi communication channels emerge at different locations of a ZigBee network, the situation will become complicated.

Figure 6 shows a possible scenario of dynamic interference in a ZigBee sensor network. In Figure 6, Wi-Fi interference source A working on Wi-Fi channel 11 is interfering the ZigBee network which is operating on ZigBee channel 23 as the frequency offset between the two systems is only 3MHz. Assume that the ZigBee PAN coordinator has sensed the interference source A in a way, it is highly possible for the coordinator to switch to

another ZigBee channel, saying ZigBee channel 17 (2435 MHz) to avoid interference. However, another interference source B is working on Wi-Fi channel 6 (2437 MHz) and locate in the vicinity of ZigBee device B. The channel switch in the ZigBee network will definitely fail since the newly chosen ZigBee channel is too close to the frequency used by interference source B. If more interference sources are introduced, the occurrence of this scenario will be highly possible, and therefore the ZigBee network channel switch mechanism will be difficult to be effective. From this motivation we propose three steps for ZigBee network to deal with such dynamic interference.

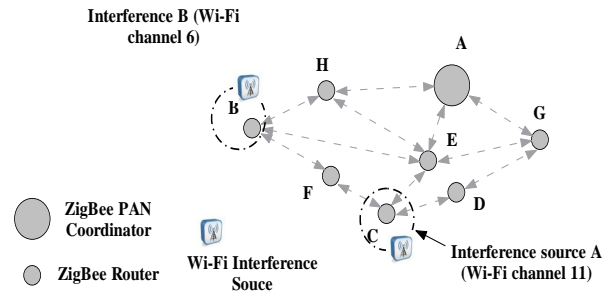


Figure 6. Multiple interference sources operating on different Wi-Fi channels

TABLE III. SUCCESS COMMUNICATION RATE OF ZIGBEE DEVICES DURING THE PERIOD OF INTERFERENCE

ZigBee Channel	Distance Y (meter)	Frequency Offset (MHz)	Interfering Energy (dBm)	Signal Strength (dBm)	Success Rate
23	2	3	-20	-47	5.26%
24	2	8	-38	-47	9.21%
25	8	13	-57	-55	89.9%
25	6	13	-57	-53	99.67%

*Step I Regular Energy Detection:* Each ZigBee router device is asked to execute energy detection on all 16 channels as a regular task. Since the ZigBee routers in a network are all static, the tolerable energy level can be obtained during the network testing. Any ZigBee router can determine which channel is available for network communication in its communication range. If all the detection results are reported to and saved on the PAN coordinator, the coordinator can have up-to-date radio environment information for later channel switch use.

*Step II Identifying the Occurrence of Interference:* The direct consequence of interference is packet loss. After sending out a data packet, a ZigBee device is unable to know if the transmission is actually successful until an acknowledgement is received from the recipient. There are two possible reasons for the sender to lose acknowledgements: a data frame is not successfully received by the recipient due to interference, or the acknowledgement frame is corrupted on its way to the sender for the same reason. Therefore, a few of settings in

the communication should help the sender to identify the occurrence of interference in the network.

When a sender is to send a data frame to a recipient, two elements should be put into the message: a unique packet sequence number and the number of retransmission which has been tried already. On the receipt of the data packet sent from the same source device, the data packet with a same sequence number will be dropped. However, the number of retransmission indicates if part of the routing path experiences a difficulty. If the number of retransmission is too high, the recipient can conclude that an interference area is possibly emerging. In detail, when the following two conditions are satisfied the recipient should inform the PAN coordinator to switch channel: 1) The number of retransmission for the same packet is over a threshold  $N_{Threshold\_Retransmission}$ , and 2) The number of different packets sent from the same sender employing retransmissions is over a threshold  $N_{Threshold\_Packet}$ .

**Step III Channel Switch:** The PAN coordinator switches channel according to the up-to-date radio environment information received from the regular energy detection implemented by all available routers. A set of simple rules are proposed here to assist alternative channel selection through the analyzing of the latest energy detection results: 1) Rule1: If the energy level of a channel is over the threshold  $E_{Threshold}$  the channel should be eliminated from available channel list.  $E_{Threshold}$  can be obtained from the network pre-testing. 2) Rule2: Remove the current channel if it remains in the available channel list after the implementation of rule 1. 3) If multiple channels remain in available channel list after rule 2 implementation, randomly select a channel as the alternative channel. 4) If there is no channel remaining in the available channel list after rule 2 implementation, select the one with a least energy level among all 16 channels.

## VI. EXPERIMENT TEST

The experimental test is for evaluating if the ZigBee network can correctly respond to dynamic interference. An experimental set-up is illustrated in Figure 7. Five ZigBee routers A, B, C, D, and E, and a coordinator form the ZigBee network. The coordinator is linked with the monitoring room. All the five routers conduct energy detection every 5 minutes and report to the coordinator.

ZigBee router A acts as a sender to send data packets to the coordinator. The ZigBee network works at ZigBee channel 15 (2425 MHz). Two Wi-Fi interference sources, W1 and W2, are located in the vicinity of ZigBee routers B and E with a 4 meter separation. Wi-Fi router W1 works at Wi-Fi channel 3 (2422 MHz) and Wi-Fi router W2 works at Wi-Fi channel 11 (2462 MHz). The distance between ZigBee router A and router C is 40 meters which is close to the maximum communication range of the ZigBee device. ZigBee router B is automatically employed in relaying the data packet from ZigBee router A to ZigBee router C as it has much stronger signal strength comparing with router C.

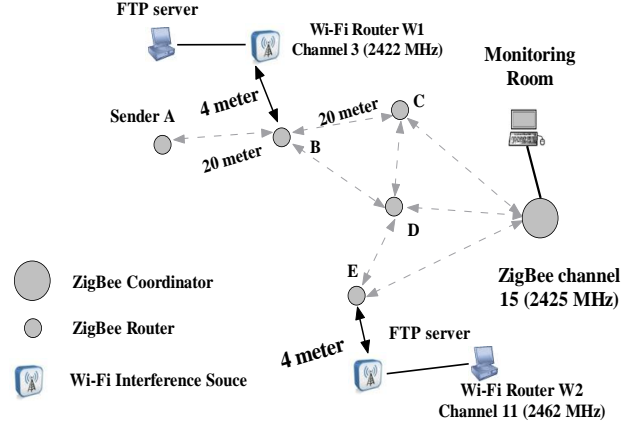


Figure 7. Experiment test and deployment

During the experiment, ZigBee router A kept sending data packet to the coordinator every 10 seconds. Meanwhile, the FTP server running on the computer connected to the Wi-Fi router W2 started to generate traffic at 700 KB/second on Wi-Fi channel 11 (2462 MHz). Because the Wi-Fi channel was 37 MHz away from the frequency used by ZigBee network (ZigBee channel 15, 2425 MHz), the Wi-Fi interference did not affect the ZigBee communication. After 10 minutes, all ZigBee routers should have reported the energy detection result to the coordinator for at least twice. The FTP server running on the computer connected to the Wi-Fi router W1 started to generate traffic at 700KB/second on Wi-Fi channel 3. The PAN coordinator detected that three different packets (i.e.  $N_{Threshold\_Packet}$ ) have employed more than 3 times retransmissions (i.e.  $N_{Threshold\_Retransmission}$ ) during half minutes. The latest energy detection results received by the coordinator show that ZigBee routers C and D reported that all channels were available to them; however, routers B and E reported that some channels were marked as bad channel. Table IV and V show the energy detection results obtained from routers B and E. The channels marked with gray color means they are unsuitable for ZigBee network communications as they have been occupied with a high energy. According to this up-to-date energy detection table, the coordinator can effectively select a suitable channel to switch. In our experiment, ZigBee channel 17 was chosen and the success transmission rate becomes 99.9%.

TABLE IV. ENERGY DETECTION RESULT FROM DEVICE B

ZigBee Channel (Centre Frequency)	11 (2405 MHz)	12 (2410 MHz)	13 (2415 MHz)	14 (2420 MHz)	15 (2425 MHz)	16 (2430 MHz)	17 (2435 MHz)
Measured Energy (dBm)	-60	-55	-38	-30	-30	-47	-61

ZigBee Channel (Centre Frequency)	18 (2440 MHz)	19 (2445 MHz)	20 (2450 MHz)	21 (2455 MHz)	22 (2460 MHz)	23 (2465 MHz)	24 (2470 MHz)	25 (2475 MHz)	26 (2480 MHz)
Measured Energy (dBm)	-61	-65	-65	-74	-76	-82	-86	-96	-98



TABLE V. ENERGY DETECTION RESULT FROM DEVICE E

ZigBee Channel (Centre Frequency)	11 (2405 MHz)	12 (2410 MHz)	13 (2415 MHz)	14 (2420 MHz)	15 (2425 MHz)	16 (2430 MHz)	17 (2435 MHz)
Measured Energy (dBm)	-96	-86	-82	-80	-80	-79	-82

18 (2440 MHz)	19 (2445 MHz)	20 (2450 MHz)	21 (2455 MHz)	22 (2460 MHz)	23 (2465 MHz)	24 (2470 MHz)	25 (2475 MHz)	26 (2480 MHz)
-74	-63	-63	<b>-32</b>	<b>-20</b>	<b>-20</b>	<b>-36</b>	-59	-61

## VII. CONCLUSIONS

The proposed interference detection and mitigation strategies in this work fully consider the interference characteristics. Through arranging network deployment, enabling dynamic interference energy detection, and setting up interference adjustment, the interference inside a ZigBee building monitoring system can be properly monitored, discovered, and responded. The contribution of this work focuses on the provision of complete and

feasible interference strategies for WSN applications in a Wi-Fi environment.

## REFERENCES

- [1] ZigBee Specification (2004), [www.zigbee.org](http://www.zigbee.org)
- [2] ZigBee Alliance, “ZigBee Enables Smart Buildings of the Future Today”, ZigBee White Paper 2007 April.
- [3] M. Petrova, J. Riihijarvi, P. Mahonen, and S. Laella, “Performance study of IEEE 802.15.4 using measurements and simulations”. Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE.
- [4] S. Y. Shin, H. S. Park, and W. H. Kwon, “Mutual interference analysis of IEEE 802.15.4 and IEEE 802.11b. Computer Networks,” The International Journal of Computer and Telecommunications Networking. Vol 51, pp. 3338–3353, August, 2007.
- [5] W. Yuan, X. Wang and J.-P. M. G. Linnartz, “A Coexistence Model of IEEE 802.15.4 and IEEE 802.11 b/g. Communications and Vehicular Technology in the Benelux,” 14th IEEE Symposium, Dec 2007.
- [6] J. Ross, “The book of Wi-Fi: install, configure, and use 802.11b wireless networking”, No Starch Press, 2003.

# Species area relations and information rich modelling of plant species variation

James Furze<sup>1\*</sup>, Quan Min Zhu<sup>1</sup>, Feng Qiao<sup>2</sup>, Jennifer Hill<sup>1</sup>

<sup>1</sup>Faculty of Environment and Technology  
University of the West of England  
Frenchay Campus, Coldharbour Lane,  
Bristol, BS16 1QY, UK  
\*email: James.Furze@uwe.ac.uk

<sup>2</sup>Faculty of Information and Control Engineering  
Shenyang Jianzhu University  
9 Hunnan East Road, Hunnan New District  
Shenyang 110168 China

## Abstract

Least squares regression is used to show the relationship of species with area on a global scale. Using a modelling based approach climatic variables are selected and made use of in a proposed information rich model of plant species variation. Future developments include advances in mathematical theory, biogeography and computer science.

*Key words: species area, modelling, mathematics, biogeography, computer science.*

## I. INTRODUCTION

The species area relationship is an inconsistently measured, inaccurate method of describing species area relations [11], hence the first aim of this study is to construct a global plant species area relationship and show how its use is becoming redundant on larger scales for estimation of species numbers due to empirically measured species numbers being recorded with non-standardized units. Furthermore, this paper introduces the use of engineering based techniques in order to demonstrate more accurate information based plant species relations, allowing the examination of larger scale trends such as plant strategies. As a result of the above, the authors will identify the relevance and development of engineering techniques in biogeography, computer science and related fields.

## II. METHOD

Calculation of species numbers was made. 20 diversity zones (DZ) were described and standardized to the number of species/10000km<sup>2</sup>. DZ 8-10 contained more than 3000 species/10000km<sup>2</sup> and these areas are investigated in the current study. Species recorded in terms of presence were sourced from the Global Biodiversity Information Facility (footnote<sup>1</sup>, [15]) in each of the DZ 8-10 locations [1]). Species numbers against

locations in latitudinal order were plotted using a histogram form of Microsoft Excel 2007©. Species Area relations were indicated by plotting species numbers with area, following the classic species area relationship:

$$S=cA^z \quad (1)$$

S= the number of species; c= a specific environmental constant; A= area in standardized units; z= constant relating to the rate of increase within the taxa present. Least squares regression ( $y=mx+b$ ) was used, where exponent z is the gradient of the line (slope m) and the intercept of the line is the logarithm of c. Species Area relations were also plotted using Microsoft Excel, 2007 ©. Statistical analysis was carried out on the regression correlation at  $p=0.05$  using SPSS, version 16.0 ©.

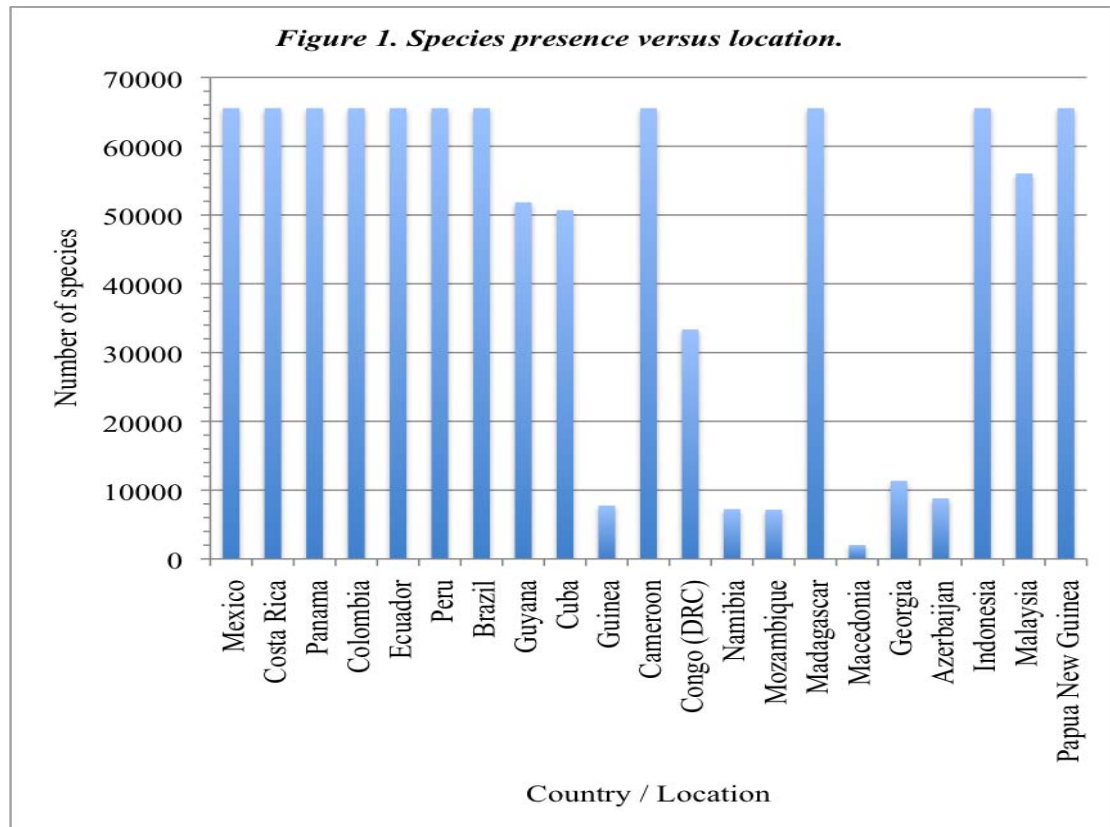
Climatic data (1960-90) were sourced from the Intergovernmental Panel on Climate Change (footnote<sup>2</sup>, IPCC) using approximate bounding longitude and latitude. Climatic parameters are described in the discussion section. Altitude data were sourced from the CIA World Factbook (see footnote<sup>3</sup>). The 7-plant strategy approach [4] was applied in an information rich mapping of species.

<sup>1</sup>Global Biodiversity Information Facility, accessed 12 10, URL: <http://www.gbif.org/>

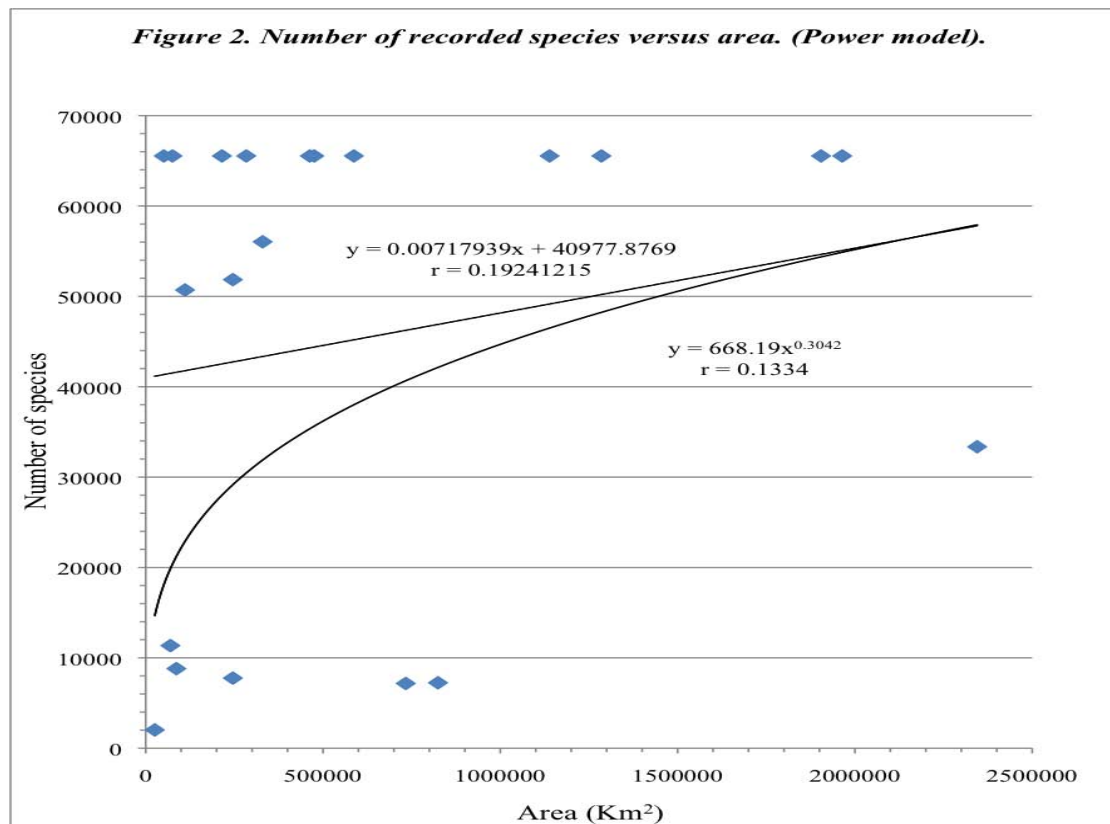
<sup>2</sup>IPCC, data distribution centre accessed 04 11, URL: <http://www.ipcc-data.org/>

<sup>3</sup>Central Intelligence Agency, world factbook, accessed 03 11 URL: <https://www.cia.gov/>

### III. RESULTS



**Figure 1.** The number of recorded species presences at the selected locations of DZ 8-10 in longitudinal order.



**Figure 2.** The number of species plotted against the area of each location in  $\text{km}^2$ .

The gradient of the straight line obtained in Figure 2 ( $m(z)$ ) is 0.00717939. The gradient of figure 2 shows a positive relationship between area and species numbers, the correlation being 0.19 (insignificant at  $p=0.05$ ). The intercept of the line ( $b$  ( $\log c$ )) is 40977.8769. The calculation for the species area relationship can be written as  $S = \log 40977.8769 A^{0.00717939}$ . A power curve is also fitted to the data to illustrate the flexible relationship.

Figure 3 a) – d) shows quarterly values of one of the variables used in the current proposed model on a global scale.

Fig. 3. a).

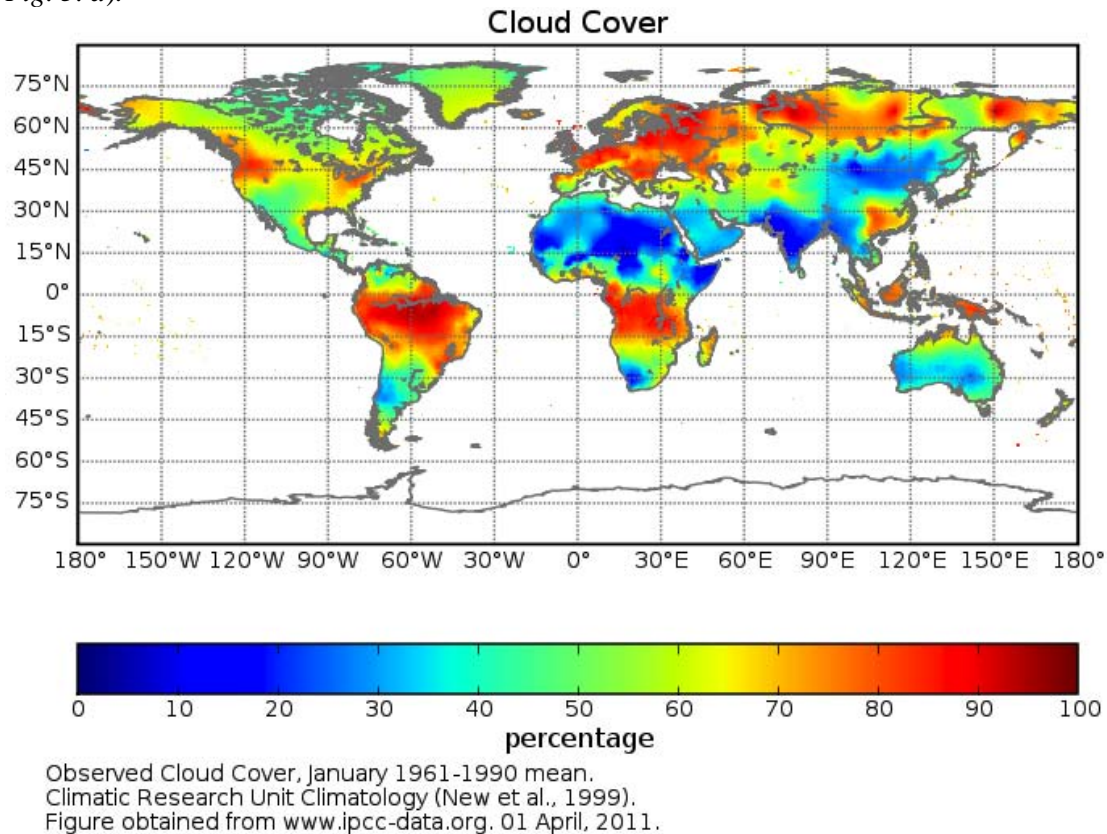


Fig. 3. b).

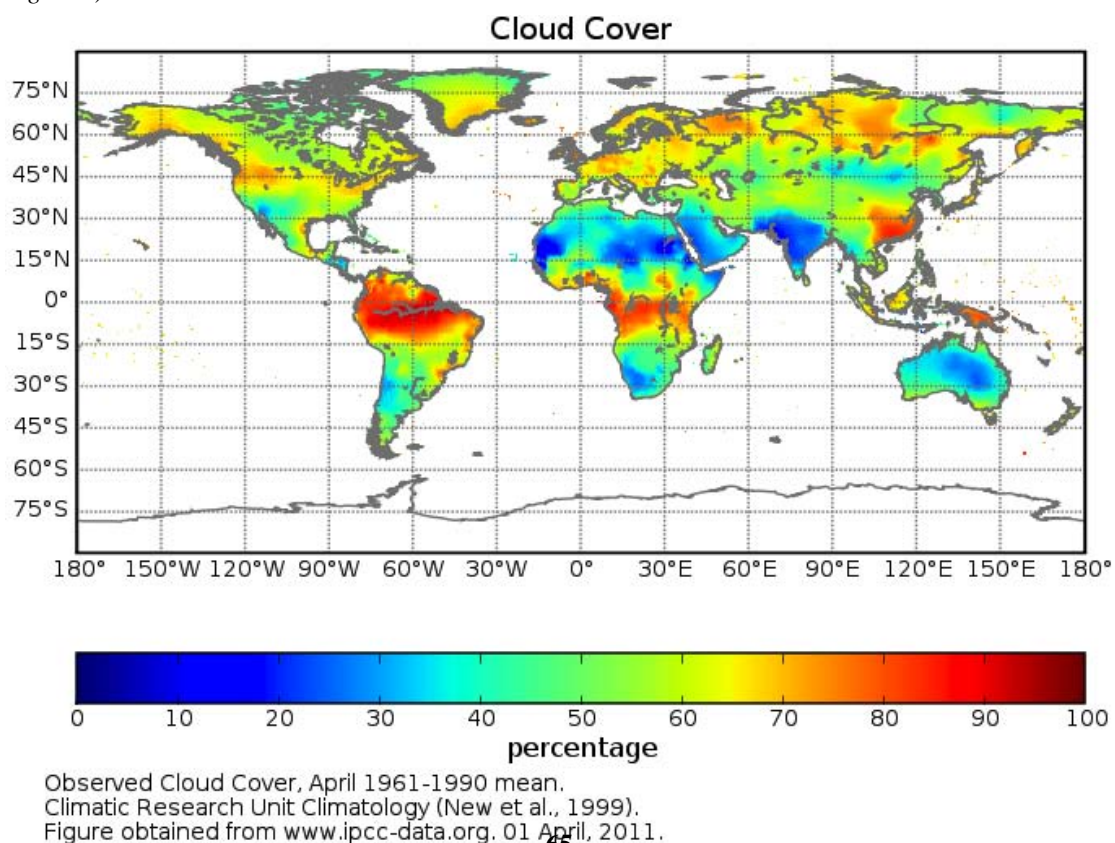




Fig 3. c).

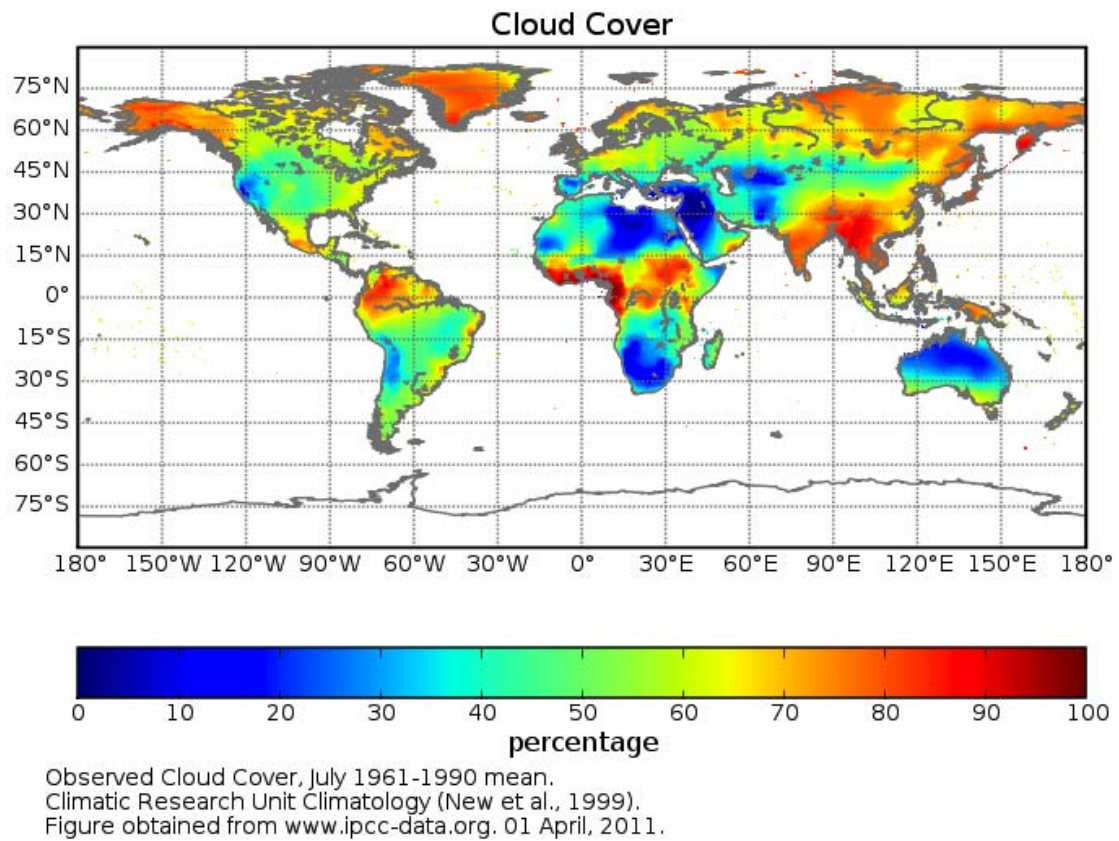
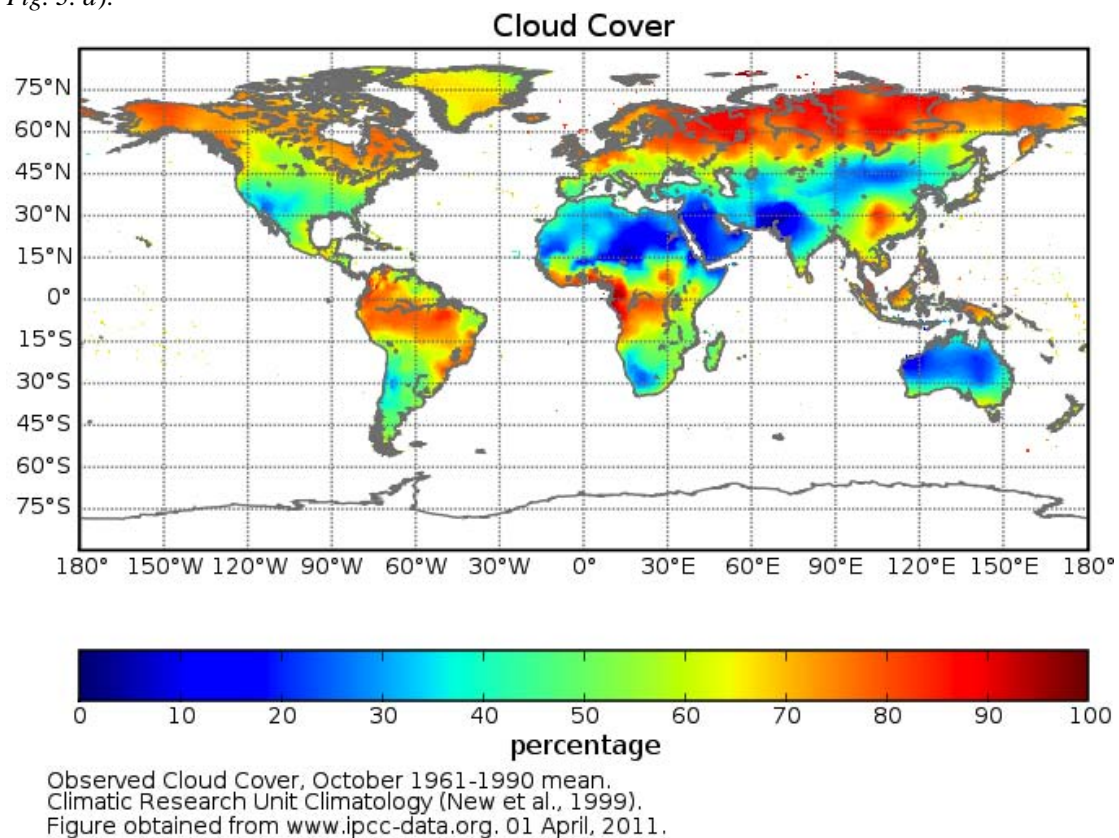


Fig. 3. d).



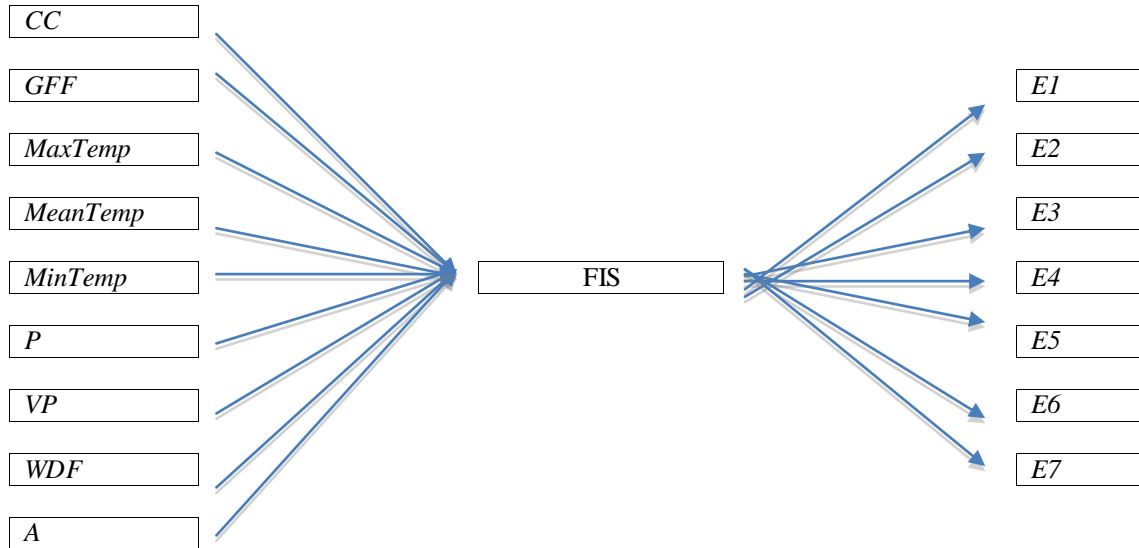
**Figure 3. Mean quarterly values of global cloud cover (1961-1990), a) January, b) April, c) July, d) October.**

Cloud cover is given as an example of a variable, which combines both elements of water and elements of energy. At this wide resolution, for example, Madagascar, within 30 and 60°East, 0-30°South shows 50-80% cloud cover in January, 40-60% in April, 40-50% in July, 50% in October.

#### IV. DISCUSSION

The correlation of species with area shown in the least squares regression of Figure 2 is insignificant (at  $p=0.05$ ). Greater levels of significance have been obtained using highly

standardized units [1]. However other factors also play a part in the relationship, as investigated [7]. The current study suggests a more information rich model to investigate patterning of species (Figure 4).



**Figure 4 basic organization of a fuzzy classification system as applied to 7 defined environments.** CC (Cloud cover), GFF (Ground-frost frequency), MaxTemp (Maximum temperature), MeanTemp (Mean temperature), MinTemp (Minimum temperature), P (Precipitation), VP (Vapour pressure), WDF (Wet day frequency) and A (Altitude) data feed into the FIS (Fuzzy inference system). The degree of membership is obtained for each of the 7 environments (E1, E2...E7).

Fuzzy logic is an appropriate system by which we develop algorithms to characterise global plant distribution as the proposed characterisation is reliant on covariates of climatic factors that are used to calculate species presence based on probability of the species occurring at particular points. Development of the current model enables us to make use of the Water-Energy based patterning of variation [5], used here to determine plant strategies [4]. For this purpose rules are generated in the FIS: for example IF P is *high* AND MeanTemp is *high* AND CC is *medium* AND A is *Low to Medium* THEN E1. The basic model is described in terms of linguistic fuzzy predicates and gives a generalised description of the conditions which best suit a strategy of plant growth. The proposed model effectively combines knowledge base and expert knowledge. This may be further refined using additional techniques. Similar use of the

techniques has been shown for environmental indices [12], to classify qualitative biological data in combination with genetic algorithm techniques [3] in marine systems [14].

Global conditions are partitioned into 7 environmental types, which also represent plant life-history based strategies. The environments are defined in terms of three main criteria for growth: the ability to compete within the environment, the ability to be tolerant to stress within the environment and the ability to reproduce in early growth successfully, with high numbers of propagules. Species are assigned a place in each of the environmental types by expert knowledge based intuition, though future work may include the use of genetic algorithms to fine-tune the characterisation. Such a framework of species organization is similar to that of [6], where the technique for order preference by similarity to ideal situation (TOPSIS) approach is combined with entropy for accurate classification. As a result of the current study, however the author suggests that additional factors may be integrated into the model to identify increased membership function [8]. Such an approach may help to elucidate further patterns such as life-forms of plant species, which are linked to plant strategy and show distinct patterns in geographic variation [4]. Furthermore, integration of soil/geological factors and increased use of species location data

may help to facilitate an algorithmic classification of plant metabolism patterns (photosynthesis) [13].

It is suggested that the validity of the methods used in this study may be verified by the use of alternative data sources as increased data become available in the future. It may be possible, for example, to measure the patterns of species presence directly by linking alternative file formats directly to software used for analysis (e.g. 'ASCII' format directly linked to MATLAB). Additionally, data obtained from such a process may include obtaining binary species presence patterns to demonstrate Euclidian proximity. Such linking may breach the gap between real and virtual based systems for use in effective modelling of the climate, though this may require application of a more complex algorithmic technique [10], [2].

## V. ACKNOWLEDGMENT

Prof. Holger Kreft for motivation for this study and Jesbin Baidya (IPCC) for data supply.

## REFERENCES

[1] Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., Kreft, H. (2005). Global Centers of Vascular Plant Diversity. *Nova Acta Leopoldina* NF, vol. 92, (342), pp. 61–83.

[2] Bornhofen, S., Barot, S., Lattaud, C. (2011). The evolution of CSR life-history strategies in a plant model with explicit physiology and architecture. *Ecological Modelling*, vol. 222, pp. 1-10.

[3] Chen, S. M., Lin, H. L. (2006). Generating weighted fuzzy rules from training instances using genetic algorithms to handle the iris data classification problem. *J. Information Science and Engineering*, vol. 22, pp. 175-188.

[4] Grime, J. P., Hodgson, J. G., Hunt, R. (1995). *The Abridged Comparative Plant Ecology*. Chapman & Hall.

[5] Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J. F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E., Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, vol. 84, (12), pp. 3105-3117.

[6] Hung, C. C., Chen, L. H. (2009). A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 13-16.

[7] Kreft, H., Jetz, W., Mutke, J., Kier, G., Barthlott, W. (2008). Global diversity of island floras from a macroecological perspective. *Ecology Letters*, vol. 11, pp. 116-127.

[8] Nasibov, E., Peker, S. (2011). Exponential membership function evaluation based on frequency. *Asian J. Math. Statist.*, vol. 4, pp. 8-20.

[9] New, M., Hulme, M., Jones, P. (1999). Representing twentieth century space-time climate variability. Part I- Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, vol. 12, pp. 829–856.

[10] Prusinkiewicz, P., Lindenmayer, A. (1990). *The Algorithmic Beauty of Plants*. Springer-Verlag, Berlin.

[11] Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

[12] Silvert, W. (2000). Fuzzy indices of environmental conditions. *Ecological Modelling*, vol.130, pp. 111-119.

[13] Su, Y., Zhu, G., Miao, Z., Feng, Q., Chang, Z. (2009). Estimation of parameters of a biochemically based model of photosynthesis using a genetic algorithm. *Plant, Cell and Environ.*, vol. 32, (12), pp.1710-1723.

[14] Taheriyoun, M., Karamouz, M., Baghvand, A. (2010). Development of an entropy-based fuzzy eutrophication index for reservoir water quality evaluation. *Iran. J. Environ. Health. Sci. Eng.*, vol. 7, (1), pp. 1-14.

[15] Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., Gray, W. A., White, R. J., Jones, A. C., Bisby, F. A., Culham, A. (2007). How global is the global biodiversity information facility? *Plos One*, vol. 11, pp. 1-10.



# Study of a Multivariable Coordinate Control for a Supercritical Power Plant Process

Omar Mohamed<sup>1</sup>, Jihong Wang<sup>2</sup>, Bushra Al-Duri<sup>3</sup>

<sup>1</sup>School of Electronic, Electrical and Computer Engineering, University of Birmingham, B15 2TT, UK

<sup>2</sup>School of Engineering, University of Warwick, Coventry CV4 7AL, UK

<sup>3</sup>School Chemical Engineering, University of Birmingham, B15 2TT, UK

[ORM808@bham.ac.uk](mailto:ORM808@bham.ac.uk), [Jihong.Wang@warwick.ac.uk](mailto:Jihong.Wang@warwick.ac.uk), [B.Alduri@bham.ac.uk](mailto:B.Alduri@bham.ac.uk)

**Abstract**—The paper presents the recent research work in study of a novel multivariable coordinate control for a 600MW supercritical (SC) power plant. The mathematical model of the plant is described in the first part of the paper. Then, a control strategy is designed based on Model Predictive Control (MPC) theory. It is noticed that the linear MPC alone performs well only within limited small load changes under a constant level of disturbances and measurement noises generalized from the prediction algorithms. So, a dynamic compensator is proposed to work in parallel with the MPC to track large load changes. Because the model has been identified with on-site closed loop response data, the multivariable optimal control signals have been used as a correction to the reference of the plant local controls instead of direct control signal applications. The simulation results show the good performance of the controller in response to the large load changes. Furthermore, it has been proved that the plant dynamic response can be improved by increasing the coal grinding capability and pulverized coal discharging through implementation of suitable coal mill controllers.

**Keywords** - *Supercritical Power Plant; Mathematical Modeling, Optimal Control of Chemical Process;*

## I. INTRODUCTION

Although there is a considerable renewable energy penetration to electrical power supply in recent years, coal fired power stations are still playing a dominant role in power generation due to its large power generation capacities. It is a worldwide big challenge for how to reduce the environmental impact brought from coal fired power generation. Supercritical boiler technology is one of the choices implemented with the improved energy efficiency. Supercritical boilers employ higher efficient Rankine cycle due to its high operating pressure and temperature (above critical points - for water 22.06 MPa and 374.15 °C pressure and temperature respectively), leading to improved thermal efficiency, lower fuel consumption and lower CO<sub>2</sub> emissions. Although supercritical and ultra-supercritical coal fired power generation is becoming the main choice of power generation technology since past two decades and will become a dominant technology on the long run, there are still some concerns about adopting this technology in the UK. It is required to fulfill the National Grid Code (NGC) requirements before adopting this technology in the UK. To satisfy the NGC, the plant should provide the primary

response to 0.5 Hz frequency deviation in ten seconds [1]. The paper is to study potential control strategies for improving SC power plant dynamic responses. The first research on optimal control of SC power plants was designed in 1978 ([2]) with identified state space model and dynamic programming for control task formulation. The application of nonlinear model based predictive control (NMBPC) was reported in [3] as preliminary results. Conventional dynamic matrix control (DMC) was published in [4] that was designed for SC power plants. However, conventional DMC may not work for the whole operating range. In [5], a model for SC unit was reported for power system frequency simulation study. The performance of diagonal recurrent neural network for predictive control of coal fired SC and ultra-supercritical (USC) power plants were shown in [6] [7] [8]. This paper aims to design a coordinated control for a 600 MW SC coal fired plant with emphasis on the target of updating the reference signals of the plant local controllers. Because the model has been identified with closed loop data, the designed control signals should be used as correction to the reference of the plant local controls (the coal mill local control, the feed-water flow control, and the turbine governing system) which is extremely helpful in power plant industries. The multivariable control system is composed of MPC in parallel cooperation with PID fully coupled structure compensator. The performances of MPC are limited within small range of once through operation, not whole range of large load changes. The use of dynamic compensator as additional means to track large steps of load changes for more reliable power plant operation. Simulation results have shown the applicability of the proposed method. It has been also proved that the coal mill local control of feeder speed and primary air fans have significant on the plant primary response of MW power output.

## II. DESCRIPTION OF THE PLANT PROCESS

For the power plant process considered here, vertical spindle mills are adopted, which is a dominant type for SC coal fired power plants ([9-10]). The raw coal enters the mill inlet tube and carries the coal to the middle of grinding rotating table. Hot primary air flows into the mill from bottom to carry the coal output from grinding process to the classifier that is a multi-stage separator

located at the top of the mill. The heavier coal particles fall down for further grinding and the pulverized coal is carried pneumatically to the furnace. Inside the boiler, the chemical energy released from combustion is converted to thermal energy. The heat is exchanged between the hot flue gas to the water through heat exchangers. The boiler contains thin tubes as heating surfaces which form the economizers (ECON), waterwall (WW), low temperature superheater (LSH), platen superheater (PSH), final stage superheater (FSH), and reheaters (RH). The water is forced at high pressure (SC pressure) inside the economizer and passes through all those heating sections. Since pressure is above the critical point, the sub-cooled water in the economizers is transferred to the supercritical steam in the superheaters without evaporation. The SC steam is then expanded through turbines. The high pressure (HP) turbine is energized by the steam supplied at the final stage superheater and the reheaters are used to reheat the exhaust steam from the HP turbine before it returns to the intermediate pressure (IP) turbine. The mechanical power is converted to electrical power by a synchronous generator coupled to the turbines. The specifications for the boiler used in this study are shown in Table I.

TABLE I. BOILER SPECIFICATION

Flow rate of superheated steam(t/h)	1728
Steam pressure (MPa)	
FSH outlet	25.1
ECON inlet	27.2
steam temperature (C°)	
FSH outlet	571
ECON inlet	277
Fuel (t/h)	Pulverized coal of 260.24

### III. MATHEMATICAL MODEL OF THE PROCESS

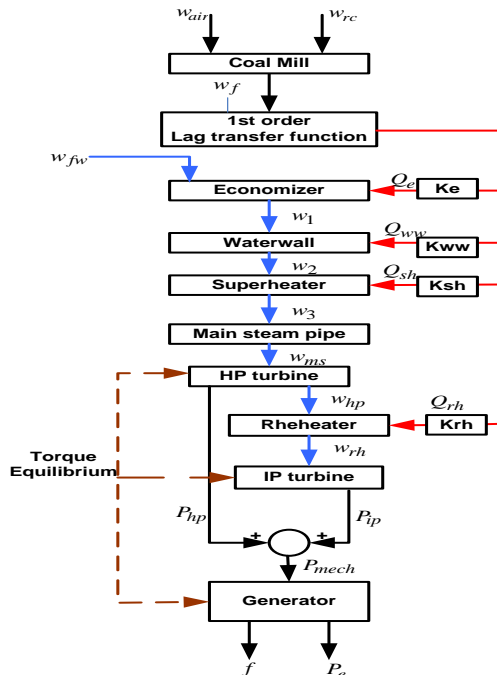


Fig.1. Illustration of a SC power plant process mathematical model

A power plant process is schematically illustrated in Fig. 1. It is a complex system with multi interconnection of subsystems and components. The first task for our study is to derive the mathematical model for the process. The detailed modeling study can be found in the previously published work by the authors [9] [10] [11]. The parameters, detailed model structure and the way how the model is derived are all described in those references. The variables appeared in Fig.1 are defined in Table II.

TABLE II. LIST OF SYMBOLS IN FIG.1

$w_{rc}$ : Raw coal flow rate (Kg/s).	$w_{ms}$ : Main steam flow rate (Kg/s)
$w_{air}$ : Primary air flow(Kg/s)	$w_{rh}$ : Reheated steam flow rate (Kg/s)
$Q_e, Q_{sh}, Q_{rh}$ , and $Q_{ww}$ : heat transferred from tube wall to the fluid (MJ/s)	$P_{mech}$ : Mechanical power (MW)
$w_f$ :Pulverized coal flow rate (Kg/s).	$f$ : frequency (p.u)
$T_{out}$ :Mill outlet temperature (C°).	$P_e$ : Electrical power (MW)
$w_{fw}$ : Feed water flow rate (Kg/s).	$w_1, w_2, w_3$ :intermediate mass flow rates (Kg/s)

The coal mill model described in [9] [10] is integrated with the rest of the plant model reported in [11]. Because there are multiple mills operating in parallel in real power plants, a proportional gain is multiplied by the pulverized coal output to obtain the total pulverized coal supplied to the burners in the plant model. The heat flow is basically related to fuel flow through proportional gains ( $K_e, K_{ww}, K_{sh}$ , and  $K_{rh}$ ) as shown in the figure with 1<sup>st</sup> order lag transfer function. For more details about modeling, please refer to [11].

Finally, the whole plant mathematical model is implemented in MATLAB/SIMULINK environment for plant response study and the new control strategies test. Identification of each subsystem parameters has been carried out using an optimization algorithm based on Genetic Algorithms (GAs) with data recorded on closed loop operation. The model has been verified over a wide operating range. Some identification and verification results are shown in Figs.3 and 4 respectively. Genetic Algorithms (GAs) is also used for identification of mill parameters. The detailed mill parameter identification is described in [9-10]. For the identification process, here is the data measured from the power plants: 1) main steam temperature; 2) main steam pressure; 3) reheater pressure; 4) steam flow rate. The measured variable data for identification of turbine /generator parameter optimization are: 1) output power; 2) system frequency. The plant data from a 600MW SC power plant is used for identification. Some of the boiler-turbine-generator unit is reported in [11].

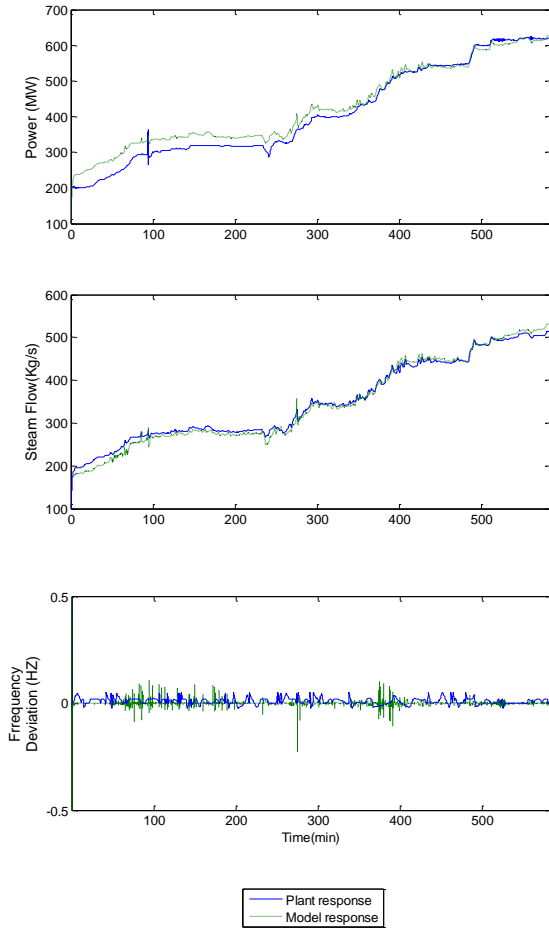


Fig.3 some identification results (power, frequency deviation, and steam flow)

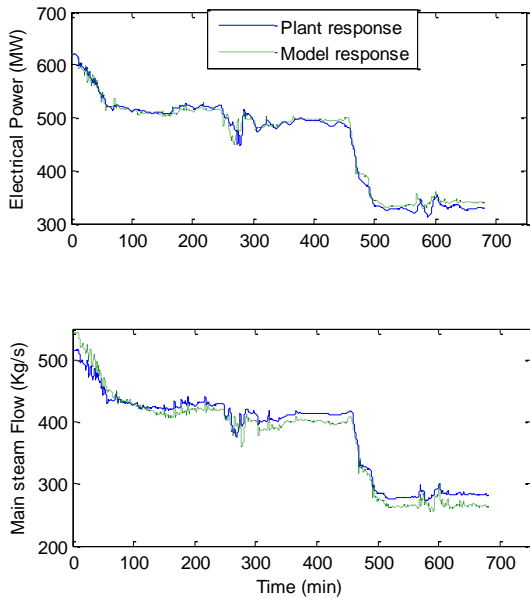


Fig.4 Some Load down verification results (Power and main steam flow)

#### IV. CONTROL SYSTEM DEVELOPMENT

The system contains many complicated nonlinear processes which require great control effort to deal with. MPC is considered as one of the well recognized control technologies for industrial process applications. MPC performs the optimization task which is based on minimizing certain cost function with pre-calculated output information and the past control moves. The complete control strategy is designed as follows.

##### A. MPC Controller Design

In this paper, the MPC tool utilizes a linear model for predicting the future output of the plant and optimizes the future plant inputs accordingly. A low order linear model that represents the main outputs to be controlled of the power plant process has been developed for MPC. The responses which have been extracted from the outputs of the nonlinear model have been used to identify and verify the linear model which has the following form:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \quad (31)$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \quad (32)$$

where  $\mathbf{x}$  is the state vector (4 states),  $\mathbf{y}$  is the output vector (3 outputs), and  $\mathbf{u}$  is the input vector (3 inputs).  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , are the normalized state space model matrices. The parameters of the digitized model are:

$$\mathbf{A} = \begin{bmatrix} 0.7759 & 0.05537 & 0 & 0 \\ 0.7615 & 0.05434 & 0 & 0 \\ 0.001564 & 0.0001116 & 0 & 0 \\ 19.14 & -23.77 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.008447 & 0.01265 & -0.3055 \\ 0.008225 & 0.01249 & -0.3334 \\ 0.02015 & 0.001941 & -0.01803 \\ 0.2067 & 0.314 & -8.378 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The inputs and the outputs which have been used for identification of the controlled plant are chosen as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \text{Main steam pressure (MPa)} \\ \text{Electrical Power (MW)} \\ \text{Main steam temperature (C}^\circ\text{)} \end{bmatrix}$$

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} \text{Feedwater flow (Kg/s)} \\ \text{Raw coal flow (Kg/s)} \\ \text{Valve Position(p.u)} \end{bmatrix}$$

The prediction of the future output with consideration of the output disturbance is formulated as follows [14]:

$$\mathbf{Y} = \mathbf{C}[\mathbf{F}\mathbf{x}(k) + \mathbf{S}\mathbf{u}(k-1) + \mathbf{\Phi}\Delta\mathbf{U}] + \mathbf{d} \quad (33)$$

where:

$$\mathbf{Y} = [y(k+1|k) \ y(k+2|k) \ \dots \ y(k+H_p|k)]^T$$

$$\Delta\mathbf{U} = [\Delta\mathbf{u}(k+1|k) \ \Delta\mathbf{u}(k+2|k) \ \dots \ \Delta\mathbf{u}(k+H_c-1|k)]^T$$

$$\mathbf{d} = [\mathbf{d}(k+1) \ \mathbf{d}(k+2) \cdots \mathbf{d}(k+H_p)]^T$$

In which  $\mathbf{y}(k+i/k)$  is the predicted output vector at instant  $(k+i)$  with the current available plant information  $k$ ,  $\mathbf{y}(k+i/k) = [y_1(k+i/k) \ y_2(k+i/k) \ y_3(k+i/k)]$  and  $\Delta \mathbf{u}(k/k) = [\Delta u_1(k/k) \ \Delta u_2(k/k) \ \Delta u_3(k/k)]$ .  $H_p$  is the prediction horizon and  $H_c$  is the control horizon.  $\mathbf{d}$  is considered as disturbance which is the estimated difference between the predicted output of the linear model and the actual output response from the nonlinear model. The matrices  $F$ ,  $S$  and  $\Phi$  are defined as follows:

$$F = \begin{bmatrix} \mathbf{A} \\ \vdots \\ \mathbf{A}^{H_c} \\ \mathbf{A}^{H_c+1} \\ \vdots \\ \mathbf{A}^{H_p} \end{bmatrix}, \quad S = \begin{bmatrix} \mathbf{B} \\ \vdots \\ \sum_{i=0}^{H_c-1} \mathbf{A}^i \mathbf{B} \\ \vdots \\ \sum_{i=0}^{H_c} \mathbf{A}^i \mathbf{B} \\ \vdots \\ \sum_{i=0}^{H_p-1} \mathbf{A}^i \mathbf{B} \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \mathbf{B} & \cdots & 0 \\ \mathbf{AB} + \mathbf{B} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{H_c-1} \mathbf{A}^i \mathbf{B} & \cdots & \mathbf{B} \\ \sum_{i=0}^{H_c} \mathbf{A}^i \mathbf{B} & \cdots & \mathbf{AB} + \mathbf{B} \\ \vdots & \vdots & \vdots \\ \sum_{i=0}^{H_p} \mathbf{A}^i \mathbf{B} & \cdots & \sum_{i=0}^{H_p-H_c} \mathbf{A}^i \mathbf{B} \end{bmatrix}$$

After predicting the future behavior for a specific horizon, the next step is to calculate the optimal control moves for the vector  $\Delta \mathbf{u}$  to minimize the objective function below:

$$\xi(k) = \sum_{i=H_w}^{H_p} \|\mathbf{y}(k+i|k) - \mathbf{r}(k+i|k)\| \mathbf{Q} + \sum_{i=0}^{H_c-1} \|\Delta \mathbf{u}(k+i|k)\|^2 \mathbf{R} \quad (4)$$

The weighting coefficient matrices ( $\mathbf{Q}$  and  $\mathbf{R}$ ), control interval ( $H_w$ ), the prediction and control horizons, these parameters are used to tune the MPC and they mainly affect the performance of the controller and computation time demands. Simulating different scenarios have been carried to decide suitable values for these parameters. The term  $\mathbf{r}$  is a vector represents the demand outputs as references for MPC. The input/output constraints are determined according to the power plant operation restrictions, which are expressed as the maximum and the minimum allowable inputs:

$$\mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max} \\ \Delta \mathbf{u}_{min} \leq \Delta \mathbf{u} \leq \Delta \mathbf{u}_{max}$$

The optimization problem is constrained optimization process to be solved by quadratic programming (QP) solver which is inherently built in MATLAB. Zero-order holder is used to convert the discrete time MPC actions to a continuous time signal for the plant. Finally, the steps for the proposed MPC are summarized as follows:

- 1- Predict the output of the plant.
- 2- Compute the errors vector for the predicted output.
- 3- Calculate the control shifts or changes  $\Delta \mathbf{u}$
- 4- Repeat the above steps for each sampling time.

#### B. The MIMO compensator design:

The plant model embeds high nonlinear features. The MPC shows good performance for limited range of once-through operation even with consideration of constant disturbance and measurement noises in prediction. The MIMO compensator is then designed by using Genetic Algorithms for wide-range load following performance. The MIMO PID with coupled structure can be designed and reported in many research articles [15] [16] [17]. The generalized form for  $n \times n$  system is then described by:

$$\mathbf{K} = \begin{bmatrix} C_{11}(s) & \cdots & C_{1n}(s) \\ \vdots & \ddots & \vdots \\ C_{n1}(s) & \cdots & C_{nn}(s) \end{bmatrix} \quad (5)$$

In which the input to the transfer function matrix is the errors vector, the output is the signal required for the compensator. Each  $s$  function in the matrix has the normalized proportional + integral + differential parameters which can be written as:

$$C_{ab}(s) = k_p + \frac{k_i}{s} + k_d s \quad (6)$$

The error between the linear and nonlinear model (in this study, the nonlinear model output represents the real power plant output) is not fixed with load variation. Therefore, the use of constant disturbance in prediction of the MPC can not provide the solutions which are powerful enough for achieving the demand performance. The use of additional loops based intelligent tuning for more robust performance is justified. Then, another performance index to be minimized is written as:

$$\varepsilon = \sum_{t=0}^N w_1 |e_P(t)| + w_2 |e_P(t)| + w_3 |e_T(t)| \quad (7)$$

The optimal control law for robust solutions becomes:

$$\mathbf{u} = \underbrace{\mathbf{u}_{mpc}}_{\text{minimizes } \xi} + \underbrace{\mathbf{u}_{co}}_{\text{minimizes } \varepsilon}$$

where  $e$  is the error remaining from plant severe nonlinearity and the subscripts to indicate the outputs responses of pressure, Power, and Temperature, respectively.  $w_1$ ,  $w_2$ ,  $w_3$  are the weighting coefficients used in optimization. In our work, the matrix  $\mathbf{K}$  is only  $3 \times 3$  matrix. The major steps performed by GA to reach the optimal solution can be found in [18]. The dynamic compensator parameters are listed in Table.III:



TABLE III. MIMO COMPENSATOR PARAMETERS

$C_{ab}$	$k_p$	$k_i$	$k_d$
$C_{11}$	2.88	0.0393	0
$C_{12}$	0.074	0	0
$C_{13}$	0.0041	0	0
$C_{21}$	0.0209	0	0
$C_{22}$	0.0054	$2.8 \times 10^{-8}$	0
$C_{23}$	0.0054	0	0
$C_{31}$	0.0862	0	0
$C_{32}$	$7.74 \times 10^{-8}$	0	0
$C_{33}$	0.0054	$7.1571 \times 10^{-8}$	0

The additional corrections to the MPC control signals are introduced. The resultant optimal control signals and their final destination are mentioned in Table IV below.

TABLE IV CONTROL SIGNALS AND THEIR SUBSYSTEMS

Optimal Control signal	Subsystem
$u_1$	Feedwater flow pump
$u_2$	Coal mill local control
$u_3$	HP turbine Control valve

There are other parallel schemes reported in the literature of chemical process control [19][20]. In [19], an adaptive neuro controller is located in parallel with MPC to increase the controller robustness while in [20] extra PID scheduled loops are installed in parallel with MPC to compensating the plant nonlinearity. The slow dynamics of the mill can be handled by its local control system. The control of vertical spindle mills is comprehensively studied. The estimated amount of pulverized coal from the mill model is compared with the optimal demand value supplied by the optimal controller that perform two control tasks in the performance indices (4), and (6), the error is used to generate control signals to the coal feeder and primary air fan. In particular, one controller is used to control the primary fan speed or power output, which tunes the pulverized coal flow rate to the furnace. Another controller is the temperature controller, which regulates the hot air flow input to control the mill temperature. The complete control package is shown in fig.6. The simulation results are presented in the next section.

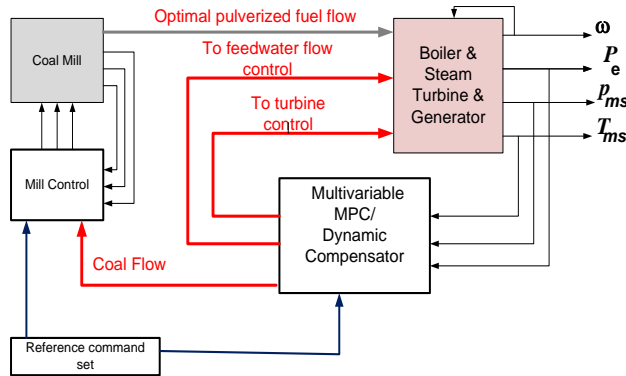


Fig.5 the proposed control strategy

## V. SIMULATION STUDY

Simulation results are mentioned in two cases to show the performance of the control strategy proposed including the effect of milling dynamics. Case. A is with the implemented local mill controllers that are tuned and prepared to receive the optimal corrected reference from the optimal control strategy while Case.B is with the existing mill control system. A step change in the load demand of 100MW is applied. The temperature setpoint was 570C° and the pressure is scheduled by look-up table of the power demand input. From the mentioned figures (Fig.6 , 7, and 8), it has been observed that the primary response of the plant power is enhanced by enhancing the coal grinding performance by following the load faster than case.B. Less pressure fluctuation is observed with the enhanced control. The mill variables are mentioned in Fig.9, higher mill pressure is created to carry more pulverized fuel to the burners, the feeder speed is increased to fill in the mill with more raw coal and keep the energy storage in the mill at desired level for giving quicker responses. This also observed in the air flow response of higher amount of air flow in Case. A. The coal mill play an important role in load following capability enhancement as expected. More water is pumped to the boiler to avoid overheating the boiler surfaces due to the increased fuel firing.

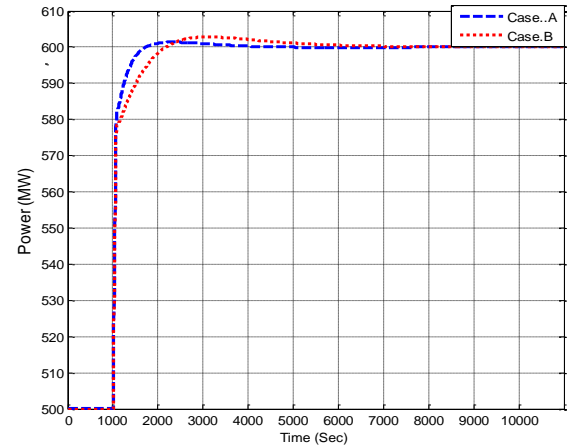


Fig.6 Electrical power dynamic response

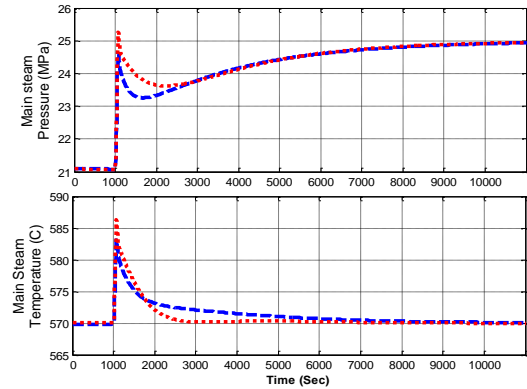


Fig.7 Pressure and Temperature

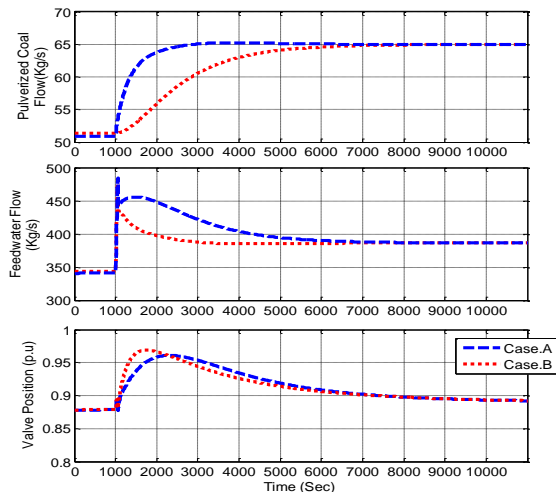


Fig.8 Manipulated Variables

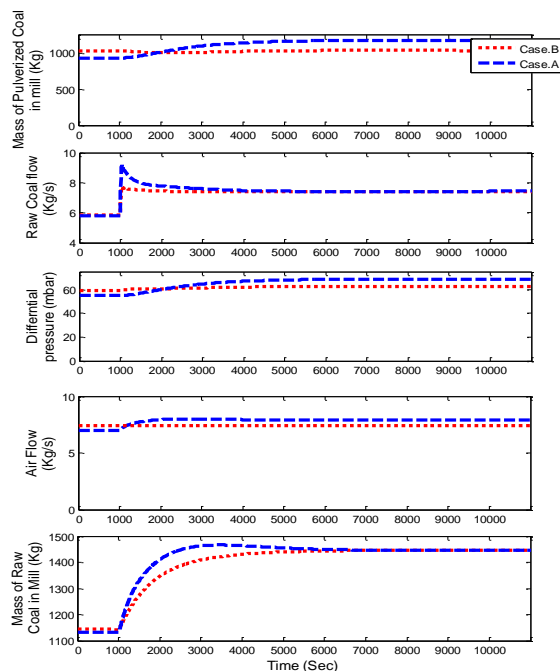


Fig.9 Variables for each mill in service

## VI. CONCLUSION

A mathematical model of a coal fired supercritical power plant is described in the paper. Simulation study indicates the model can predict the main dynamic response variation trends of the 600MW SC power plant process. A coordinate control is proposed and tested through simulation study with relatively large load variations. The control considers the milling process as a key stage for power plant dynamic response improvement. From the study, it is convinced that better control of the milling process can improve the power plant load following capability and/or primary response of the plant power.

## ACKNOWLEDGEMENT

These authors would like to thank UK EPSRC (grant EP/G062889/1) and AWM/ERDF Birmingham Science City Energy Efficiency and Demand Reduction project for their funding support.

## REFERENCES

- [1] Grid (UK), The Grid Code, Issue 4 Revision.5, 2010. Available at <http://www.nationalgrid.com/uk/Electricity/Codes/gridcode/gridcode/docs/> on 14<sup>th</sup> March 2011.
- [2] H. Nakamura, H. Akaike "Statistical Identification for Optimal Control for Supercritical Thermal Power Plants", *Automatica*. Vol.17, No.1, pp 143-155, 1981.
- [3] B.P. Gibbs, D.S. Weber, D.W. Porter, "Application of Nonlinear Model Predictive Control to Fossil Power Plants", proceedings of 30th conference on decision and control, Dec 1991. pp.1850-1856.
- [4] J. A. Rovnak, R. Corlis, "Dynamic Matrix Based Control of Fossil Power Plant", *IEEE Transactions on Energy conversions*. Vol. 6, No. 2, pp. 320-326, 1991.
- [5] T. Inoue, H. Taniguchi, Y. Ikeguchi "A Model of Fossil Fueled Plant with Once-through Boiler for Power System Frequency Simulation Studies", *IEEE Transactions on Power Systems*, vol. 15, No. 4, pp1322-1328, 2000.
- [6] J. Kwang, Y. Lee, J.S. Heo, J. A. Hoffman, S-H Kim, and W-H Jung, "Neural Network Based Modeling for Large Scale Power Plant", *Power Engineering Society General Meeting*, 2007. IEEE Volume, Issue 24-28, June 2007, pp.1-8.
- [7] K. Lee, J.S. Heo, J.A. Hoffman, S-H. Kim, W-H. Jung, "Modified Predictive Optimal Control Using Neural Network-based Combined Model for Large-Scale Power Plants", *Power Engineering Society General meeting*, June 2007, pp. 1-8.
- [8] K. Lee, J. H. Van Sickle, J. A. Hoffman, W-H. Jung, and S-H. Kim, "Controller Design for Large-Scale Ultrasupercritical Once-through Boiler Power Plant", *IEEE Transaction on Energy Conversions*, vol.25, No.4, pp1063-1070, 2010.
- [9] J. Wei, J. Wang, and Q. H. Wu, "Development of Multisegment Coal Mill Model using and Evolutionary Computation Technique", *IEEE Transaction on Energy Conversions*. vol. 22, pp. 718-727, 2007.
- [10] Y.G. Zhang, Q.H. Wu, J. Wang, G. Oluwanda, D. Matts, D., and X.X. Zhou, "Coal mill modelling by machine learning based on on-site measurement", *IEEE Transactions on Energy Conversion*, Vol.17. No.4, pp549-555, 2002.
- [11] Omar Mohamed, Jihong Wang, Bushra Al-Duri "Mathematical Modeling of Coal Fired Supercritical Power Plant and Model Parameter Identification Using Genetic Algorithms". Lecture notes in Electrical Engineering, "Electrical Engineering and Applied Computing", Chapter.1. Springer. July. 2011
- [12] K. Rayaprolu, "Boilers for Power and Process" CRC Press, 2009
- [13] Yao-Nan. Yu "Electric Power System Dynamics" Academic Press, 1983.
- [14] J.M.Maciejowski "Predictive Control with Constraints". Prentice Hall. 2001.
- [15] Alberto Herreros, Enrique Baeyens, Jose' R. Pera'n "Design of PID- type controllers using multiobjective genetic algorithms". *ISA Transaction*, 41, 2002. pp. 457-472.
- [16] Wei-Der Chang "A multi-crossover genetic approach to multivariable tuning PID controllers tuning". *Expert Systems with Applications* 33 (2007) 620-626.
- [17] K. Y. Lee, Robert Demio "Boiler-Turbine Control System Design Using Genetic Algorithms". *IEEE Transaction on Energy Conversions*, Vol.10, No.4, 1995.
- [18] D. Goldberg "Genetic Algorithms in Search, Optimization, and Machine Learning". Addison Wesley. 1989.
- [19] Po-Feng Tsaia, Ji-Zheng Chub, Shi-Shang Janga, Shyan-Shu Shieh "Developing a robust model predictive control architecture through regional knowledge analysis of artificial neural networks". *Journal of Process Control*. *Journal of Process Control* 13 (2002) 423-435.
- [20] H. Peng, W. Gui, K.Nakano, and H. Shiya "Robust MPC Based on Multivariable RBF-ARX Model for Nonlinear Systems", *Proceedings of IEEE Conference on Decision and Control*, Seville, Spain, 12-15, Dec. 2005.

# Robust Stability and performance with $H_2/H_\infty/\mu$ controller for Single Person Aircraft

J. Mashayekhi Fard<sup>1</sup>, M.A. Nekoui<sup>2</sup>, A. Khaki Sedigh<sup>3</sup>, R. Amjadifard<sup>3</sup>

<sup>1</sup> Department of Electrical Engineering, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran

<sup>2</sup> Department of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran

<sup>3</sup> Department of Engineering Faculty, Tarbiat Moallem University, Hesarak, Tehran, Iran

( E-mail: [mashayekhi@iaus.ac.ir](mailto:mashayekhi@iaus.ac.ir) , [manekoui@eetd.kntu.ac.ir](mailto:manekoui@eetd.kntu.ac.ir) , [sedigh@kntu.ac.ir](mailto:sedigh@kntu.ac.ir) , [amjadifard@tmu.ac.ir](mailto:amjadifard@tmu.ac.ir) )

**Abstract**— In a physical system several targets are normally being considered in which each of nominal and robust performance has their own strengths and weaknesses. Nominal performance means considering system operation without uncertainty, has decisive effect on the operation of system. Robust performance means considering operation with uncertainty. This target causes intensive limitation on the controller grade and even it's solutionless. The target of this paper is to present a new approach for balancing between nominal and robust performance, using their own gains. This will be done, using one new approach. First, the controller of  $H_2/H_\infty$  will be designed for nominal performance target, robust stability and noise reduction and then  $\mu$  controller is designed for robust performance. By combining these two controllers and achieve their weights. Finally, Simulation and comparison studies are used to show the effectiveness and benefits of method.

**Keywords**-  $H_2/H_\infty/\mu$  Controller, Multi-Objective Control, Single Person Aircraft, LMI.

## I. INTRODUCTION

Un modeled dynamic, non linearity of systems and the availability of disturbances cause that linear control systems theory never reaches the ideal solution. For this, several targets are attended in a system control. 1) Robust Stability: meaning that system will be stable with uncertainty. 2) .nominal performance: without considering uncertainty, the fault of system minimizes. 3) Robust performance: with considering uncertainty, the fault of system minimizes. To consider robust performance, use  $\mu$  analysis. Usually, the availability of robust performance causes to extreme limitation on controller and dominating it from reaching feasible condition, and in case of achieved feasible condition, it increase the grade of controller, and the resulted controlling signal increases and causes to saturate actuator. But in some of the systems it will be used, compulsorily. For example, in power systems which transient response is decisive, robust performance will be considered. 4) Operating limitation on controlling signal: increase of controlling signal causes saturation of actuators.  $H_2$  norm essence can be responsible for such target. 5) Minimizing disturbance effect: distortion can cause undesirable effect of transient response, so reducing the effect of disturbance, is one of the controlling targets. Mixed norm of  $H_2$  and  $H_\infty$  can be a useful strategy, to reach noted controlling targets. To date, number of studies has been performed on the mixed

norm and multi- objective control. In this paper the target is implementation of approach on multi – inputs controlling systems. One of the useful on multi- inputs controlling system is the Optimal state feedback. In such kind of systems, feedback matrix is not unique, and abundance is available to select the release grade. This degree of freedom can be used to designing K for system robustness. Prevalently, it is possible to design a controller which regularly includes 5 noted targets above. In a more explicit description, controller includes two parts, the first one using mixed  $H_2$  and  $H_\infty$  norm and the other using  $\mu$  synthesis. These two parts , include weights each of which have Important roles in systems control, because each of 2 and 3 targets , has its own definiteness which their combination can create a new solution . In operation, we look for minimizing faults. If the available error function is not desirable, using a suitable weight function can lead us to the target. So, designing weight function is extremely important. At first, a controlling problem will be changed to LFT standard form, considering uncertainty, then status equations will be written and using constraint's weight function will be determined to reach the robust controlling targets. Then, using robust state feedback method, now it is time to select them again , using state feedback and weight functions repetition methods to supply the robust performance of system in an acceptable way. The first formulation of  $H_\infty$  control problem performed in 1981 by Zames. Next to Zames, Doyle, Zhou, and Glover were premiers of robust control, To date, vast number of researches have performed studies in robust control,  $H_2$  control and  $H_\infty$  control. Doyle et al in [1] analyzed the state space with  $H_\infty$  and  $H_2$  standard form and its solving. The conditions of solving problem and its solution using Hamiltonian matrix introduction are of importance of this paper. This paper is a comprehensive reference that has been worked in many other research works. Doyle et al in [3] present a tutorial overview of linear fractional transformations (LFTs) and the role of the structured singular value,  $\mu$ , and linear matrix inequalities (LMIs) in solving LFT problems. They focus on two standard notions of robust stability and performance,  $\mu$  stability and performance and Q stability and performance, and their relationship. Doyle et al in [4] present A tutorial introduction to the complex structured singular value ( $\mu$ ) is presented, with an emphasis on the mathematical aspects of  $\mu$ , The  $\mu$  -based methods discussed here have been useful for analyzing the

performance and robustness properties of linear feedback systems. Lescher et al in [6], designed multivariable, multi-objective controller to set wind turbine. Controlling problem of this paper is minimization of  $H_2/H_\infty$ . LMI solves this problem. Rotea et al in [8] combined  $H_2/H_\infty$ . Two important approaches are presented. 1)  $H_2$  optimized control with  $H_\infty$  bound (in fact a bounded optimization). 2) Simultaneous  $H_2/H_\infty$  optimized control. In each step, problem formulation and controller were performed. Patton et al in [11], presents a theoretical and tutorial treatment of the eigenstructure assignment approach for robust fault detection. Liu et al in [12] presented a new approach for robust control design of multivariable systems via eigenstructure assignment, genetic algorithms and gradient-based optimization. Doyle et al in [16] are shown that different of a mixed  $H_2$  and  $H_\infty$  infinity norm arise from different assumptions on the input signals. Akbar et al in [19], study a mixed  $H_2/H_\infty$  control law is derived using auxiliary cost minimization approach for continuous time linear time invariant singularly Perturbed System (SPS). The rest of this paper is as follows. Section 2 the description of aircraft model and control system. Section 3 establishes the problem to be addressed, the  $H_2/H_\infty$ ,  $\mu$  and  $H_2/H_\infty$ ,  $\mu$  combination control will be demonstrated. In Section 4, illustrates the approach and the results of simulations will be discussed. Section 5 ends the paper.

## II. AIRCRAFT MODEL

In an airplane five main sections could be listed as: motor, body section, landing system and wheels, wing and tail. The pitch angle of an airplane is controlled by adjusting the angle (and therefore the lift force) of the rear elevator. The aerodynamic forces (lift and drag) as well as the airplane's inertia are taken into account. The X-29 aircraft is a recent example of a control configured vehicle that was designed with a high degree of longitudinal static instability (up to 35 percent at low subsonic speeds). The vehicle is stabilized by a full-authority, fly-by-wire flight control system. Linear models were used extensively prior to flight to determine the close loop stability, controllability, and handling qualities with the various control system modes through the flight envelope. This section describes the commercial aircraft models which is implemented. In [21] which is a comprehensive report of NASA, it has been researched over X-29 state equation and model. In [22] has been designed only the  $H_\infty$  controller over X-29. The X-29 airplane is a relatively small, single seat, high-performance aircraft powered by a single F404-GE-400 engine (General Electric, Lynn, Massachusetts). Empty weight is 6350 kg. The vehicle incorporates a forward-swept wing with close-coupled canards to provide a low-drag configuration. The airplane physical characteristics are presented in table 1. The aircraft model is obtained by linearizing the nonlinear equations of motion about a 280 ft/sec (307Km/hr) landing configuration [21]. The three input three output model

which describes the longitudinal dynamics is given as follows: [21-22]

Table 1 X-29 physical characteristics

Maximum trust force	8130 N.M	Horizontal speed	$v$
Angel attack	$\alpha$	Canard area	$3.437 m^2$
Strake flap position	$\delta_{sf}$	Symmetric flap position	$\delta_{sf}$
Pitch Euler angel	$\theta$	Pitch rate	$\dot{\theta}$
Wing span	8.29m	Wing area	$17.185 m^2$
Canard position	$\delta_c$	weight	6350 kg

$$\dot{x} = Ax + Bu, y = Cx, x = \begin{bmatrix} v - (\text{ft/sec}) \\ \alpha - (\text{rad}) \\ \dot{\theta} - (\text{rad/sec}) \\ \theta - (\text{rad}) \end{bmatrix}, u = \begin{bmatrix} \delta_c - (\text{deg}) \\ \delta_{sf} - (\text{deg}) \\ \delta_{sf} - (\text{deg}) \end{bmatrix}, y = \begin{bmatrix} \theta - (\text{rad}) \\ v - (\text{ft/sec}) \\ \alpha - (\text{rad}) \end{bmatrix} \quad (1)$$

$$A = \begin{bmatrix} 0.0427 & -8.5410 & 0.4451 & -32.16 \\ 0.0008 & 0.5291 & 0.9896 & 0 \\ 0.0004 & 3.5420 & 0.2228 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 0.0338 & 0.0939 & 0.0049 \\ 0.001 & 0.0013 & 0.0004 \\ 0.0272 & 0.0057 & 0.0135 \\ 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & 0 & 57.3 \\ 1 & 0 & 0 & 0 \\ 0 & 57.3 & 0 & 0 \end{bmatrix}$$

## III. PROBLEM STATEMENT

### A. $H_2/H_\infty$ Controller

Existence of uncertainty which is created due to an uncertain and erratic input (noise and disturbance) and Un-modeled dynamic is caused which we can not describe completely and precisely a true system by a mathematical model at all. On the other hand, in a true system are important the following objects that are: robust stability, robust and nominal performance, settling time, maximum over shoot and etc which try to gain these objectives about the controlling problem [19], [6], [8]. Type of uncertainty is an important in analysis. In [16] researched in optimization approach of mixed  $H_2$  and  $H_\infty$  norm.

**Theorem 1** (small gain Theorem) suppose  $M \in RH_\infty$  and let  $\gamma > 0$ , then the interconnected system shown in figure 1 is well-posed and internally stable for all  $\Delta(s) \in RH_\infty$  with  $\|\Delta\|_\infty < \gamma^{-1}$  if and only if  $\|M\|_\infty < \gamma$  [18].

Additive uncertainty shown in figure 2 robust stability task is:

$$q = (I + KG)^{-1} KP \Rightarrow \|(I + KG)^{-1} K\|_{H_\infty} < \gamma^{-1} \quad (2).$$

The objective for the inner loop control is to design a state feedback law such that the close loop system satisfies that following performance specifications,

**Objective 1:** if  $\Delta = 0$  then  $\|FS\|_\infty < 1$  (nominal performance).  $S = (I + GK)^{-1}$  ( $S$  is sensitivity function and  $F(s)$  is weighting function).

**Objective 2:** if  $\Delta \neq 0$  then system has been robust stability.

$$M = (I + KG)^{-1}K, \text{ if } \bar{\sigma}(\Delta(j\omega)) \leq \gamma(j\omega) \Rightarrow \|\gamma(S)M\|_{\infty} < 1$$

**Objective 3:**  $n$  is white noise with one PSD (power spectral density).  $H_2$  Norm, cause decreasing of controlling signal.  $\|T_{nU_1}\|_{H_2} < 1$  (To minimize U1 variance with noise input).

From parseval equation and objective 3:

$$\|Y\|_2^2 = \int_0^{\infty} Y^T(t)Y(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y^*(j\omega)Y(j\omega)d\omega$$

$$Y(j\omega) = G(j\omega)U(j\omega) \Rightarrow$$

$$\|Y\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} U^*(j\omega)G^*(j\omega)G(j\omega)U(j\omega)d\omega$$

$$\leq \frac{1}{2\pi} \sup \bar{\sigma}(G^*(j\omega)G(j\omega)) \int_{-\infty}^{\infty} U^*(j\omega)U(j\omega)d\omega =$$

$$\|G(s)\|_{\infty}^2 \|U(s)\|_2^2 \Rightarrow \sup \frac{\|Y\|_2^2}{\|U\|_2^2} = \|G(s)\|_{\infty}^2 \rightarrow$$

$$\frac{\|U_1\|_2}{\|n\|_2} \leq \|T_{nU_1}\|_{\infty} < 1$$

Then we have three tasks for controller design ( $\|FS\|_{\infty} < 1, \|\gamma(S)M\|_{\infty} < 1, \|T_{nU_1}\|_{\infty} < 1$ ), such that,

$$\left\| \begin{bmatrix} FS(K, G) \\ \gamma M(K, G) \\ RT((K, G)) \end{bmatrix} \right\|_{\infty} < 1 \quad (3). \text{ Problem (3) shown in figure 3.}$$

Rotea and doyle offer two method for solution of this problem [8],[16]. A large class of system with uncertainty can treat as LFT (Linear fractional Transformation). LFT model shown in figure 3.  $W$ : the disturbance signals to the system which won't be a function of states of system,  $Z$ : the variable that will be controlled,  $P$ : the nominal open loop system,  $Y$ : the system measurable output. To transform the changed diagram of figure 4 to the LFT model, will be write the problem to standard form, and will be dissolve using of ricati equation. The (2) LFT model is practicable in form (4) and can design a controller by theorem 2. State space of figure 3 written in (4). Determining 3 weight matrices, specified in fig.4, contain special importance. Using robust Optimal state feedback for (4) equations, will be configured next to determining controller, they will be reselected using weight functions repetition and State feedback methods, to supply the system operation, acceptably. And this is a new approach.

$$\begin{aligned} \dot{x} &= Ax + B_1 W + B_2 u \\ z &= C_1 x + D_{11} W + D_{12} u, \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} FS \\ \gamma M \\ RT \end{bmatrix} r \\ y &= C_2 x + D_{21} W + D_{22} u \end{aligned} \quad (4)$$

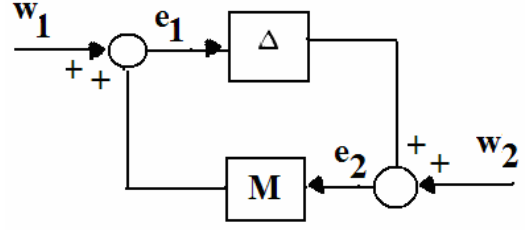


Fig 1.  $M - \Delta$  Model

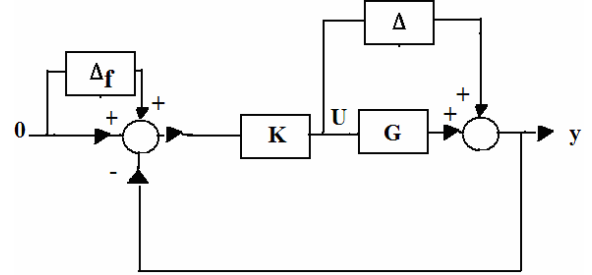


Fig2. Additive uncertainty

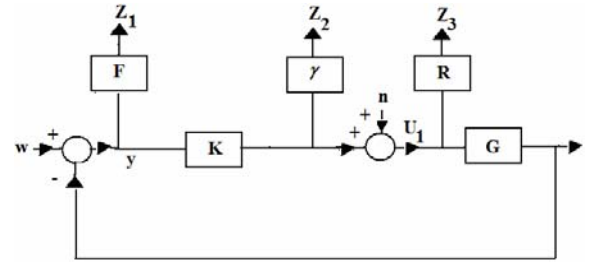


Fig 3. Graphical model of  $H_2/H_{\infty}, \mu$  combination problem with additive uncertainty

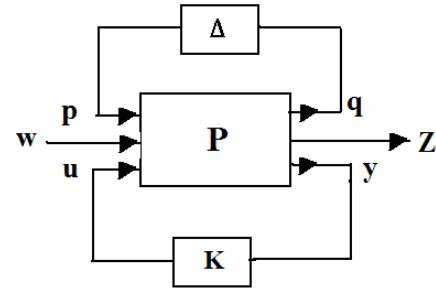


Fig 4. LFT Model

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{x}_f \\ \dot{x}_\gamma \\ \dot{x}_R \end{bmatrix} &= \underbrace{\begin{bmatrix} A & 0 & 0 & 0 \\ -B_f C & A_f & 0 & 0 \\ 0 & 0 & A_\gamma & 0 \\ 0 & 0 & 0 & A_R \end{bmatrix}}_{A_{CL}} \begin{bmatrix} x \\ x_f \\ x_\gamma \\ x_R \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & B & B \\ B_f & -B_f D & -B_f D \\ 0 & 0 & B_\gamma \\ 0 & B_R & B_R \end{bmatrix}}_{\begin{bmatrix} B_1 & B_2 \end{bmatrix}} \begin{bmatrix} r \\ n \\ u \end{bmatrix} \\ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ y \end{bmatrix} &= \underbrace{\begin{bmatrix} -D_f C & C_f & 0 & 0 \\ 0 & 0 & C_\gamma & 0 \\ 0 & 0 & 0 & C_R \\ -C & 0 & 0 & 0 \end{bmatrix}}_{\begin{bmatrix} C_1 \\ C_2 \end{bmatrix}} \begin{bmatrix} x \\ x_f \\ x_\gamma \\ x_R \end{bmatrix} + \underbrace{\begin{bmatrix} D_f & -D_f D & -D_f D \\ 0 & 0 & D_\gamma \\ 0 & D_R & D_R \\ 1 & -D & -D \end{bmatrix}}_{D_{CL}} \begin{bmatrix} r & n & u \end{bmatrix} \end{aligned} \quad (5)$$



### B. $\mu$ Controller:

Here we try to assess robust performance of this closed-loop system using  $\mu$ -analysis associated. Robust performance condition is equivalent to the following structured singular value  $\mu$  test [3].

$$\|T_{wz}(M, \Delta)\|_{\infty} < \gamma^{-1} \forall \|\Delta\|_{\infty} < \gamma \Leftrightarrow \mu_{\Delta P}(M) < \gamma \quad \forall W \quad (6)$$

The complex structured singular value  $\mu_{\Delta(M)}$  is defined

$$\text{as } \mu_{\Delta(M)} = \frac{1}{\min \left\{ \bar{\sigma}(\Delta) \mid \det(I - M \Delta) = 0 \right\}} \text{ Lower and}$$

Upper bond of  $\mu$  can be shown to  $P(UM) \leq \mu_{\Delta}(M) < \min \bar{\sigma}(DMD^{-1})$ .

### C. New approach: $H_2/H_{\infty}$ , $\mu$ combination

Now, we tend to synthesize two collectors according to figure.5. As mentioned before, the availability of robust performance causes extreme limitation on the controller, which sometimes prevents it from reaching a possible condition. Also, availability of nominal performance means considering operation without uncertainty, and it is usual that the essence of uncertainty has decisive effect on the operation. So, we tend to balance between robust and nominal performance.  $W_1$  and  $W_2$  are weight functions. Having this data, we can determine which frequencies have more uncertainty effect, with the consideration, the controller effect of  $\mu$ . Of course, it is of importance to mention that robust performance contains nominal performance, so, controller coefficient of  $\mu$  should be smaller than  $H_2/H_{\infty}$  controller coefficient.

**Problem 3-3-1:** Determine  $W_1$  and  $W_2$ , in a way that a additive uncertainty system contains robust stability.

$$M = (W_1 K_1 G + W_2 K_2 G + I)^{-1} (W_1 K_1 + W_2 K_2) \quad (7)$$

$$\|M\|_{\infty} < 1$$

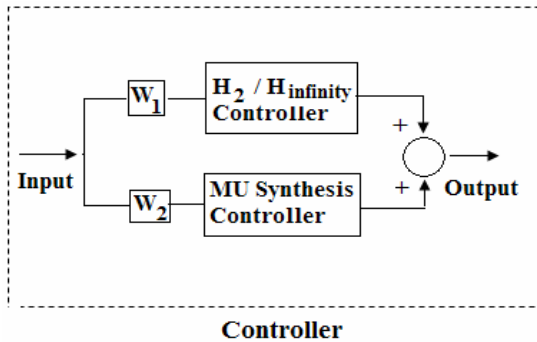


Figure 5 Controller  $H_2 / H_{\infty} / \mu$

**Problem 3-3-2:** Determine  $W_1$  and  $W_2$ , in a way that a system having multiplication uncertainty contains robust stability.

$$M = (GW_1 K_1 + GW_2 K_2 + I)^{-1} (GW_1 K_1 + GW_2 K_2) \quad (8)$$

$$\|M\|_{\infty} < 1$$

We use of state space to solve the problem 3-3-1 and 2-3-2.

#### 1) Robust optimal state feedback

We now attempt to follow the analysis of the conditioning of the pole placement problem with the multi-input case which is called the generalized state feedback. State feedback matrix not being unique in the MIMO system, there are many degrees of freedom (DoF) in this choice. These degrees of freedom are used for the following purposes: 1) positioning of the eigenvalues and the associated eigenvectors, 2) Designing  $K$  (gain feedback matrix) for low cost solution. 3) Designing  $K$  for system robustness. In this paper we assume that the state of the generalized plant  $G$  is available for feedback. To be more precise let a state-space description of  $P$  (figure 4) is given by (LFT Model):

$$\begin{aligned} \dot{x} &= AX + BU + B_W W \\ Y &= X \\ U &= KX \\ \dot{x} &= (A + BW_1 K_1 + BW_2 K_2)X + B_W W \end{aligned} \quad (9)$$

The signal  $W$  denote disturbance. The signals  $U$  and  $Y$  denote the control input and the measured output, respectively. Next to gaining  $K_1$  by  $H_2/H_{\infty}$  and  $K_2$  by  $\mu$  analysis, we tend to determine weight functions, using linear matrix inequality.

**Lemma1** (bounded-real lemma) given a constant  $\gamma > 0$ , for system,  $M(s) = (A, B, C)$  the following two statements are equivalent, 1. This system is stable  $\|M(s)\|_{\infty} < \gamma$ , 2. There exists a symmetric positive definite matrix  $Q$ , such that:[24]

$$\begin{bmatrix} A^T Q + Q A & Q B_p & C_q^T \\ B_p^T Q & -\gamma^{-1} I & D_q^T \\ C_q & D_q & -\gamma^{-1} I \end{bmatrix} < 0$$

$$Q > 0 \quad (10)$$

LMI of system (8), considering BRL theorem will be (11). ( $\beta = \gamma^{-1}$ )

$$\begin{bmatrix} (A + BW_1 K_1 + BW_2 K_2)^T Q + Q(A + BW_1 K_1 + BW_2 K_2) & Q B_W C^T \\ B_W^T Q & -\beta I & D^T \\ C & D & -\beta I \end{bmatrix} < 0 \quad (11)$$

$Q > 0$ ,  $W_1 > 0$ ,  $W_2 > 0$

We multiply the  $\begin{bmatrix} Q^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$  on Left and right of the matrix, Define

$$Y_1 = W_1 K_1 \Omega, Y_2 = W_2 K_2 \Omega, \Omega = Q^{-1}$$

Substituting into (11) yields:

$$\begin{bmatrix} A\Omega + BY_1 + BY_2 + Y_1^T B + Y_2^T B + \Omega A^T & B_W & \Omega C^T \\ B_W & -\beta I & D^T \\ C\Omega & D & -\beta I \end{bmatrix} < 0$$

$$\Omega > 0, Y_1 > 0, Y_2 > 0$$

$$W_1 = Y_1 \Omega^{-1} K_1^{-1}, W_2 = Y_2 \Omega^{-1} K_2^{-1} \quad (12)$$

## 2) METHODOLOGY

- To design the H2 /H $\infty$  for the process with uncertainty. (It helps to select the weighting function properly).
- for H2 /H $\infty$  design can use rotea and doyle method

$$([8],[15],[16]) \text{ or use } \left\| \begin{bmatrix} FS(K, G) \\ \gamma M(K, G) \\ RT((K, G)) \end{bmatrix} \right\|_{\infty} < 1 \text{ and obtained}$$

K1.

- for  $F, \gamma$  and  $R$  we use Automatic Weight Selection Algorithm[23].

d. To design the  $\mu$  controller for the process with uncertainty (if the process is unstable, at first must be stabilize). D-K iteration method can be used to improve the performance of the controller design for the system. Peak value of the  $\mu$  (D-K iteration) bound should be less than one, and obtained K2

e. Order reduction method can be used to reduce the order of the I+GK, and transfer to state space equation and given  $A_c$ .

f.  $K_1 = B^{-1} \times (A - A_{c1})$ ,  $K_2 = B^{-1} \times (A - A_{c2})$ .  $A, B$  are in P-K system (LFT Model).

g.  $W_1, W_2$  are given with LMI (11) then The robust stability of the system has to be established.

h. H infinity norm of  $W_2$  must be smaller than  $W_1$

i.  $K = W_1 K_1 + W_2 K_2$ .

this controller (k) has robust stability and desired performance.

## IV. SIMULATION RESULTS

The longitudinal dynamics of an aircraft one natural mode: the short period mode. For the X-29, the short period mode, however, consist of stable mode and an unstable mode.

At first we design H2 /H $\infty$  controller and design  $\mu$  controller. These controllers designed for P-K system. This system has four input four outputs. Then we use residualization method for reducing the order of a I+GK. By considering to practical experiments and in according to (2) weight functions selection with

$$R = \frac{0.0001(S + 0.01)}{S + 1} I, F = \frac{0.2222S + 0.6667}{0.11S + 1} I,$$

$$\gamma = \frac{0.0001(0.25S + 2)}{1.5S + 0.0001} I. K_1 \text{ and } K_2 \text{ Design with equations}$$

2 and 8.  $W_1, W_2$ , Are obtained by equation 14. In according to figure 5, K is design. K be given by:

$$K = 1.0e+002 * \begin{bmatrix} 0.0002225567461 & -0.12281728088088 & 0.21572124485198 & -1.55614052040448 \\ -0.03378163163132 & -0.00017600806331 & 0.00025499961793 & -0.00004193410595 \\ 0.00579665158345 & 0.00002881776966 & -0.00004165700284 & -0.00000554058893 \\ 0.06559002889283 & 0.00033440535085 & -0.00048241399165 & -0.00000459046638 \end{bmatrix}$$

$$W_1 = 100 \times \begin{bmatrix} -0.00126343724096 & 2.86913660898690 & -3.55125379906492 & 4.67639646437599 \\ 0.10974375219439 & 0.80143218199220 & -0.59133738035974 & 1.09521927373565 \\ -0.01883105399660 & -0.13749201248564 & 0.10143643203533 & -0.18789095696525 \\ -0.21307608469859 & -1.55587636407115 & 1.14792286477763 & -2.12619009731752 \end{bmatrix}$$

$$W_2 = 0.1 \times \begin{bmatrix} 0.00000000232699 & -0.00000164702740 & 0.00004595837565 & -1.89421994688728 \\ -0.00000035365344 & -0.00000012579903 & -0.00000028519857 & 0.00006178748615 \\ 0.00000006071014 & 0.00000002163676 & 0.00000004845379 & 0.00000743821153 \\ 0.00000068702279 & 0.00000024449670 & 0.00000055130763 & -0.00001383601001 \end{bmatrix}$$

at first we design the weight functions and were drawn in fig.6, a, b, c. By considering to practical experiments, are selected the weighting functions [23]. The singular values of LFT model are drawn in fig.6-d.

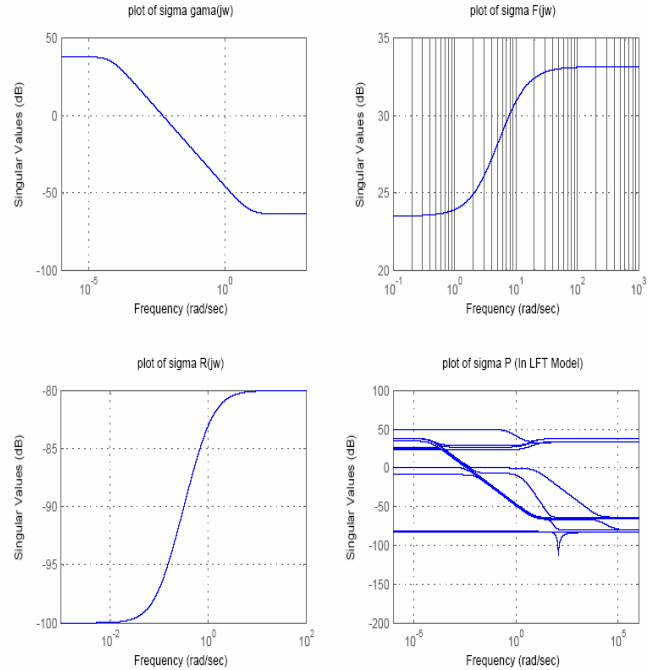


Fig 6. Singular value for weighting functions and P (LFT Model)

The Step response of T function for LFT model shown in fig.7. Note that, as regards, the system is multi output-input, the weight and sensitivity functions are the shape of matrix. The sign of success is the combination of nominal and robust performance, together. Reaching the targets with the minimum controlling signal is of the gains of noted controller. The open-loop system is unstable, but the close-loop system has appropriate results.

## V. CONCLUSION:

Each of nominal and robust performance has their own strengths and weaknesses. Nominal performance means considering system operation without uncertainty has decisive effect on the operation of system. Robust performance means considering operation with uncertainty. It is obvious that whatever the singular



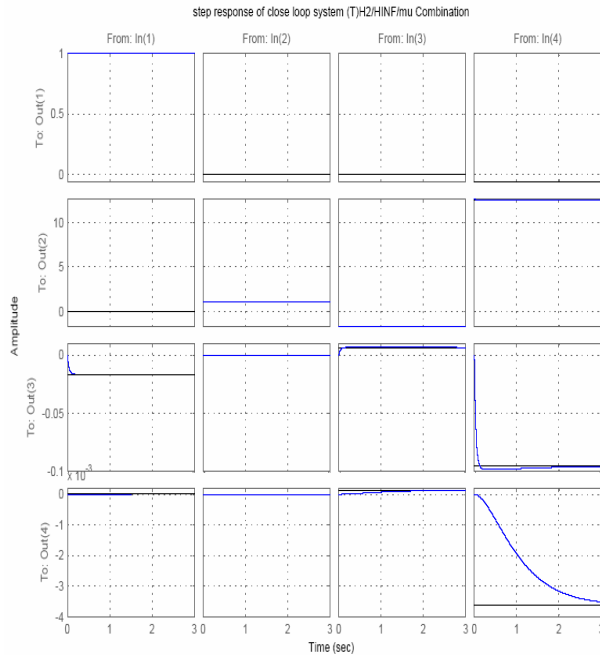


Fig 7 Step responses for T sensitivity function

values of controller are higher, systems performance is more desirable, but, a problem is created to being saturate actuators. Using the  $H_2$  norm, can avoid over-increasing the controlling signal. To design the filters or otherwise the weighting function, also, has an important role to determine the answers of close-loop. This paper tends to reduce controlling signal, robust performance, and robust stability and to design weight functions. One new approach of present paper is the combination of two controllers of  $\mu$  and  $H_2/H_\infty$ . The controller for robust stability status, nominal performance, robust performance and noise deletion are designed continuously. At first, the controller of  $H_2/H_\infty$  will be designed for nominal performance targets, robust stability and noise reduction and then  $\mu$  controller for robust performance. Now, add up these two controllers and achieve their weights. Controller will be achieved from solving optimization problem. Because of multivariable exceptional values system, controller and considered inputs-outputs were drawn. Using one low pass filter and two high pass filters, we tended to optimize the solutions. First, has been written X-29 air craft state space equations. Then, the controller has been designed. The drawn figures results indicate that this unstable system, has become stable by presence of uncertainty, and has appropriate desired performance.

## VI. REFERENCES

- [1] J.C.Doyle, K. Glover, P.P.Khargonekar, State-Space Solution to Standard  $H_2$  and  $H_\infty$  Control Problems, IEEE Transactions on Automatic Control, VOL. 34, NO. 8, AUGUST 1989.
- [2] P. Huang and K. Zhou, "Robust Stability and Performance of Uncertain Delay systems with Structured Uncertainties," IEEE Conference on Decision and Control, pp. 1509-1514, Sydney, Australia, 2000.
- [3] J. C. Doyle, A. Packard, and K. Zhou, "Review of LFTs, LMIs, and  $\mu$ ," Proc. IEEE Conference on Decision and Control, pp. 1227-1232, Brighton, UK, Dec. 1991.
- [4] A.Packard, J. Doyle, The Complex Structured Singular Value, Automatica, Vol. 29, No. 1, pp. 71-109, 1993.
- [5] L.H.Keel, S. P. Bhattacharyya, Robust, Fragile, or Optimal? IEEE Transactions on Automatic Control, VOL. 42, NO. 8, AUGUST 1997.
- [6] F.Lescher, J.Yun Zhao, A.Martinez, Multiobjective  $H_2/H_\infty$  Control of a Pitch Regulated Wind Turbine for Mechanical Load Reduction, European Wind Energy Conference, Athens, Greece, 2006.
- [7] C.Scherer, The Riccati Inequality and State-Space  $H_\infty$  - Optimal Control, PHD Thesis, UniversitÄt WÄurzburg, 1990.
- [8] M.A. Rotea, P.P. Khargonekar,  $H_2$ -optimal Control with an  $H_\infty$  constraint The State Feedback Case, Automatica, Vol. 27, No. 2, pp. 307-316, 1991.
- [9] C. Scherer, P.Gahinet and M. Chilali, Multiobjective Output-Feedback Control via LMI Optimization, IEEE Transactions on Automatic Control, VOL. 42, NO. 7, JULY 1997.
- [10] W.Tan, Zhan Xu, Robust analysis and design of load frequency controller for power systems, Electric Power Systems Research 79, 846-853, 2009.
- [11] R.J. Patton and J. Chen. On eigenstructure assignment for robust fault diagnosis. Int. J. Robust and Nonlinear Control, 10(14):1193-1208, 2000.
- [12] G.P.Liu, R.J. Patton, Robust control design using eigenstructure assignment and multiobjective optimisation, Int. Journal of Systems Science, vol. 27, no. 9, pp. 871-879, 1996.
- [13] W. Beaven, M.T. Wright and D.R. Seaward, "Weighting Function selection in the  $H_\infty$  design process", control eng. practice, vol. 4, no. 5, pp. 625-633, 1996
- [14] H.Liang, F.Yan-Ming, G.R.Duan, Multiobjective control synthesis based on parametric eigenstructure assignment, Control, Automation, Robotics and Vision Conference, Kunmlng, China, 2004.
- [15] K. Zhou, K. Glover, B. Bodenheimer, J.C. Doyle. Mixed  $H_2$  and  $H_\infty$  performance objectives. I: robust performance analysis. IEEE Trans. Automat. Control, vol. 39,no. 8, 1564-1587, 1994.
- [16] J.C. Doyle, K. Zhou, K. Glover, B. Bodenheimer. Mixed  $H_2$  and  $H_\infty$  performance objectives. II. Optimal control. IEEE Trans. Automat. Control, , vol. 39,no. 8, 1575-1587, 1994.
- [17] A.Lanzon, Weight Selection in Robust Control: An Optimisation Approach, Wolfson College, Control Group, Department of Engineering, University of Cambridge A dissertation submitted for the degree of Doctor of Philosophy, October 2000.
- [18] V. Raissi Dehkordi, B.Boulet, Frequency-Domain Robust Performance Condition for Controller Uncertainty in SISO LTI Systems: A Geometric Approach, Journal of Control Science and Engineering, Hindawi Publishing Corporation, Article ID 746762. Volume 2009.
- [19] S.A. Akbar, A.K. Singh, K.B. Datta, Study of Response and Robustness Measures of Mixed  $H_2/H_\infty$ , LQG and  $H_\infty$  Controllers for Continuous-Time Singularly Perturbed Systems, Faculty of Electrical Engineering Universiti Teknologi Malaysia, VOL. 11, NO. 2.
- [20] Zhou, Doyle, Essential Robust Control, Prentice hall, 1998.
- [21] J.Bosworth, Linearized aerodynamic and control law models of the X-29 airplane and comparison with flight data, NASA technical memorandum 4356,1992.
- [22] Y.K.Tae, development of an interactive modeling, simulation, animation and real time control aircraft environment, MSC Thesis, Arizona state university, 2000.
- [23] Sarath S Nair, Automatic Weight Selection Algorithm for Designing H Infinity controller for Active Magnetic Bearing, International Journal of Engineering Science and Technology (IJEST), VOL3,NO 1, 2011.

# Grid Roadmap based Real time Path Planning

M. R. B. Bahar, H. B. Bahar & F. Hashemzadeh

Department of Electrical Engineering & Computer Engineering

Tabriz University

Tabriz, IRAN

[mrbahar@live.com](mailto:mrbahar@live.com), [hbbahar@tabrizu.ac.ir](mailto:hbbahar@tabrizu.ac.ir) & [hashemzadeh@tabrizu.ac.ir](mailto:hashemzadeh@tabrizu.ac.ir)

**Abstract**— *The probabilistic roadmap (PRM) is forceful for path planning in static environments. Also PRM based methods may be employed for real time path planning in dynamic environments. These methods for real time path planning desire vast amount of time for preprocessing. To mitigate desired time for initialization, we propose a new method based on Grid Roadmap (GRM) which is an edge less roadmap. By suggested roadmap and utilizing a training method for robot manipulator, we attain a shape deformed model for obstacles that cancels our ambition for configuration space. Accordingly, grid roadmap construction will be on workspace. Finally, the planner searches for a collision free path in workspace with dynamic and shape changing obstacles.*

**Keywords**- *Path planning, Grid Roadmap, Dynamic environment, Shape deformed model.*

## I. INTRODUCTION

Nowadays robots are operating in areas where there may be different obstacles and humans. To satisfy a safe working area, path planning indeed will be an essential part of robots programming. Path planning requires the creation of optimized path which avoids static and dynamic obstacles in a given workspace [1].

The PRM path planning method employs probabilistic node generation in configuration space (C-Space) [2]. This stochastic method may cause to face a free narrow area of nodes or accumulation of nodes around a particular area. In addition, neighbor nodes searching and establishment of collision free edges may lead to even more iteration and time for processing. Besides, real time path planning methods based on PRM such as Dynamic Roadmap (DRM) [3], includes steps where the nodes and edges are generated in configuration space and then mapped to workspace cells. PRM generation and mapping require enormous processing time [4]. To alleviate PRM preprocessing time we should take considerations to mitigate waste time required steps. In this situation, edge segmentation and collision checking seems to require most operation time. Attention should be taken to have another basis path planning method instead of PRM that generates an alternate simple edge less roadmap. Consequently, we will require reduced amount of time for initialization. In addition, most of path planning methods desire to maneuver on configuration space and finally the resultant path is mapped to workspace. Mapping from workspace to configuration space and then vice versa takes considerable amount of time, which is not appropriate for path planning in

dynamic environments. Presented problems motivate us to look for a path planning method which is free of edges to improve path planning execution rate and operates on workspace instead of configuration space.

In this article, we propose Grid Roadmap (GRM) method for path planning which is free of edges. Also the GRM creation is on workspace with no desire for configuration space to improve path planning required time. In this work, robot manipulator trains on GRM to extract morphed model for obstacles. Training will create mapping matrix that deforms the obstacles shape and invalidates the corresponding roadmap cells. For real time path planning, obstacle morph matrix brings obstacles deformed shape and invalidates collide cells of roadmap. Then local planner searches for a collision free path on the workspace from current position to destination.

The remainder of the article is organized as in the following. In section 2 we introduce the Grid Roadmap and obstacles shape deforming technique. Section 3 describes planner. In section 4 we discuss about manipulator locating and obstacles distinction on GRM. In section 5 we implement proposed method on robot manipulator in dynamic environment. Section 6 compares GRM and DRM path planning methods. Finally, conclusions are drawn in section 7.

## II. GRID ROADMAP

As we know, PRM path planning has a robust performance in static environments and also is an excellent basis for more fast lazy PRM path planning [5] and real time path planning methods such as DRM [3]. First we take a brief review on DRM method to have comparisons by our proposed method in following steps. DRM uses a pre-computed workspace mapping for fast invalidation of blocked roadmap parts. Each workspace grid cell stores a list of roadmap nodes and edges that are in collision with the cell [4]. After initialization that takes lengthy time, each obstacle on workspace invalidates relevant edges and nodes and then local planner searches for a collision free path from current position to destination on free edges and nodes. Dynamic Roadmap (DRM) preprocessing steps [3] are:

1. Generation of PRM in a free of obstacle C-space.
2. Mapping PRM edges and nodes to workspace cells.

Subsequent to DRM initialization, in real time path planning, two steps below execute until achieving destination [3]:

3. Invalidating edges and nodes via workspace obstacles.
4. Searching for a path from current position to destination on valid edges and nodes.

Steps 1 and 2 (initialization steps) require establishment of edges and nodes and then mapping them to workspace cells. Edges creation and mapping them in workspace take vast preprocessing time [4]. To reduce path planning initialization time, we propose a new method that plans an edge less roadmap with uniform distribution of cells and operates on workspace with no desire for C-Space. We call our method Grid Roadmap (GRM). GRM consists of two main sections; Initialization and planning. After preprocessing step, path planning is executed on real time. GRM initialization for a robot manipulator is described in the following subsection:

#### A. GRM INITIALIZATION

GRM preprocessing is described in two steps below:

1. We employ an arrange cells in workspace (GRM). These cells information is stored in a same dimension matrix called mapping matrix. Mapping matrix is 4D matrixes for 2D workspace and 6D matrixes for 3D workspace with all zero indices. Mapping matrix at last creates obstacles de-shaped form.
2. We situate manipulator on each cell of GRM. Next we distinct robot arm and assign nearest cells to arm. We utilize  $\infty$ -norm in workspace as distance function declared by (1).

$$D_{\infty}^w(q, q') = \max_n \|q' - q\| \quad (1)$$

The nearest cells to arm for a particular location of manipulator are defined by (2).

$$\text{Near Cells} = \{q \mid \forall q', D_{\infty}^w(q, q') \leq ucl\} \quad (2)$$

Where  $q$  is  $(i, j)$  index of GRM, and  $q'$  is all arm segments in workspace.  $ucl$  Or unit cell length is defined by (3) for 2D workspace.

$$ucl = \max \left( \frac{\text{Workspace lenght}}{\text{Mapping matrix columns}}, \frac{\text{Workspace height}}{\text{Mapping matrix rows}} \right) \quad (3)$$

Also  $ucl$  is defined by (4) for 3D workspace.

$$ucl = \max \left( \frac{\text{Workspace lenght}}{\text{Mapping matrix columns}}, \frac{\text{Workspace height}}{\text{Mapping matrix rows}}, \frac{\text{Workspace width}}{\text{Mapping matrix depth}} \right) \quad (4)$$

An example of 40\*20 GRM (Right half plane) on 2D workspace and 2-DOF robot arm are shown in Figure 1. The end effector of arm is located on (1.7, 1.8) and near cells to arm's elbow are designated by blue circles.

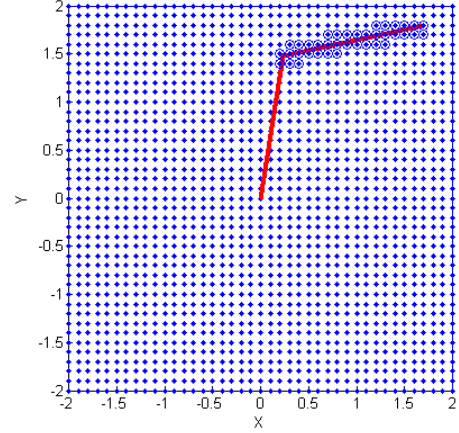


Figure 1. 2-DOF Robot manipulator on 40\*20 GRM.

For Figure 1 example  $ucl$  is 0.1, links length are 1.5 and we consider 150 segments on elbow. As seen in Figure 1, obstacle zone must be out of Near Cells and colliding Near Cells will invalidate end effector position. We define Near Cells index  $(i, j)$  and end effector index  $(k, l)$ . Then mapping matrix  $(i, j, k, l)$  index is set to one. And this step is repeated for all GRM cells to assign all Mapping matrix indices.

In the following subsection we clarify application of mapping matrix for invalidating workspace cells via obstacles shape deforming. Note that shape deforming is a potent tool for visual based robot control [6] and character animating [7, 8].

#### B. OBSTACLES SHAPE DEFORMOING

Subsequent to Mapping matrix construction, we have the invalidate cells for a particular obstacle in position  $(i, j)$  on GRM. For  $(i, j)$  obstacle index, any  $(k, l)$  that sets  $(i, j, k, l)$  index of mapping matrix to one illustrates invalid cell index on GRM. Consequently, for  $(i, j)$  obstacle index on GRM, invalid cells indices on GRM are defined by (5).

$$\text{Invalid Cells} = \{(k, l) \mid \exists (i, j), \text{Mapping matrix}(i, j, k, l) = 1\} \quad (5)$$

The designed examples illustrated in Figure 2 indicates an obstacle location on (1.4, 1) and (1.5, -0.1) and invalidate cells are designated by red pluses for a 2-DOF arm.



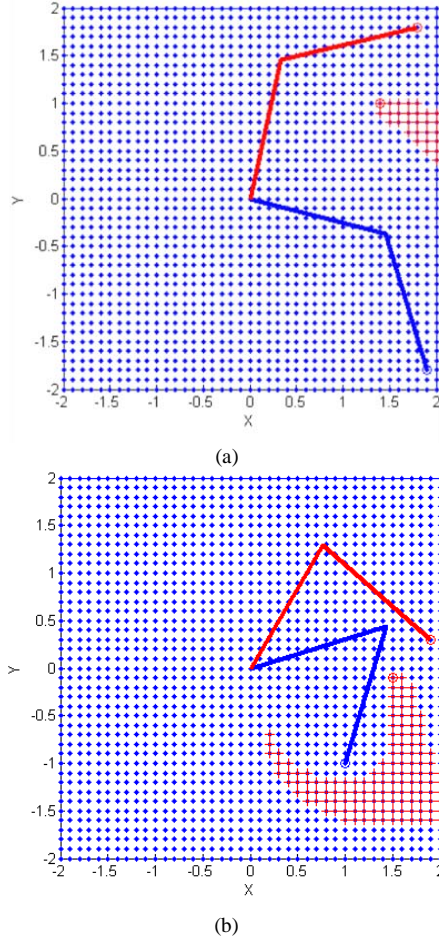


Figure 2. A spot obstacle and counterpart invalid cells. (a) Obstacle on location (1.4, 1). (b) Obstacle on location (1.5, -0.1).

A simple meaning of obstacle deformed shape is that if a spot obstacle locates on (1.4, 1) or (1.5, -0.1) as shown in Figure 2, the manipulator end effector is unable to establish on invalid cells (red pluses). Finally a path is searched from start to end on GRM cells by avoiding invalid cells.

### III. ATTRACTING AND REPELLING PLANNER

We use an attracting and repelling planner similar to potential fields [9, 10] on GRM. The near cells to destination are more weighted to attract manipulator end effector and distant cells are less weighted. Invalid cells are assigned an immense negative weight, so no path will generate on invalid cells. Near cells to invalid cells are assigned negative weights and outlying cells are all less negative weights and then the effect of destination and invalid cells are adjoin together to generate Cost matrix. A path is searched from start to end on more positive weighted indices on Cost matrix. Cost matrix dimension for Figure 2 example are 40\*20 as the same GRM dimensions. For weight determination of Cost matrix we utilize 2-norm in workspace as distance function stated by Equation 6.

$$D_2^w(q, q') = \|q' - q\| = \left( \sum_{i=1}^n (q'_i - q_i)^2 \right)^{\frac{1}{2}} \quad (6)$$

Each cell of GRM is able to bond to eight adjacent cells although neglecting marginal cells. The planner starts planning from current position and explores for more positive weighted near cells on Cost matrix. The planner connects current position to more weighted near cell and makes it present cell. And again searches for more positive weighted cells around it and repeats searching until achieving destination. A generated path for Figure 2 (a) example which starts from (1.9, -1.8) to (1.8, 1.8) and avoiding invalid cells is illustrated in Figure 3.

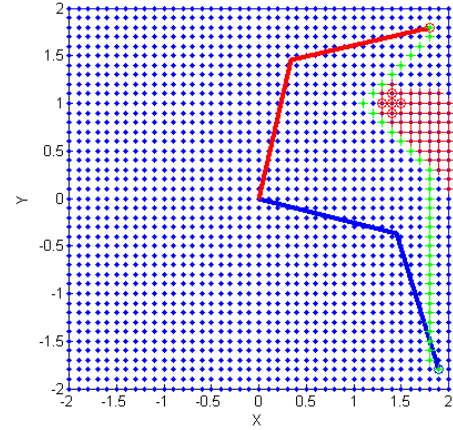


Figure 3. Generated collision free path for static spot obstacle

To make a safe area for path planning we enhance obstacle and add four pseudo spot obstacles on restrictions of original obstacle, And give more attraction gain to destination. These tasks are not crucial for static obstacles but may be applicant for dynamic environments.

A vital problem of this method is eschewing local minima's and escaping them. Hence, probabilistic methods or any other methods may be employed [11].

### IV. MANIPULATOR LOCATING AND OBSTACLE DISTINCTION ON GRM

As we divide workspace to many discrete cells, manipulator start and end position and obstacle location may not be exactly on cells. For this reason we connect manipulator end effector start and end position to nearest cells by 2-norm distance function declared by Equation 6 and then planner searches for a collision free path from nearest cell to start trough nearest cell to destination. In addition, we distinct obstacle to smaller segments and enhance each obstacle segment to 4 near cells. The enhanced shape of obstacle is defined by Equation 7.

$$\text{Enhanced Obstacle} = \{q \mid \forall q', D_2^w(q, q') \leq ucl\} \quad (7)$$

Where enhanced obstacles cells are  $q$  and GRM cells are  $q'$ . We utilize Enhanced Obstacle shape instead of original obstacle in all progresses of path planning. Addition to this distinction, for dynamic path planning we add pseudo boundary obstacles to Enhanced Obstacle for more safe area in real time path planning.

## V. 2-DOF ROBOT ARM PATH PLANNING IN DYNAMIC ENVIRONMENT

We design an example for dynamic path planning for a 2-DOF robot manipulator which starts from (1.9, -1.8) to (1.8, 1.8) location. A spot obstacle is moving from (1.4, 1) to (1.4, 1.2) during path planning. Note that we add 4 pseudo spot obstacles on boundaries of original obstacle. Planner creates collision free path for each manipulator moving steps. Results are illustrated in Figure 4.

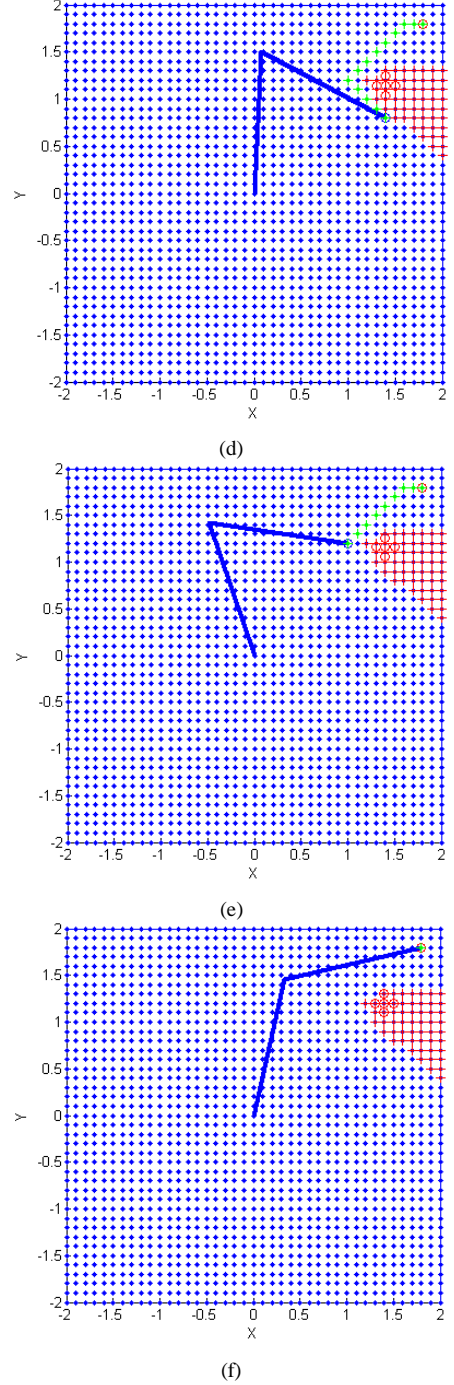
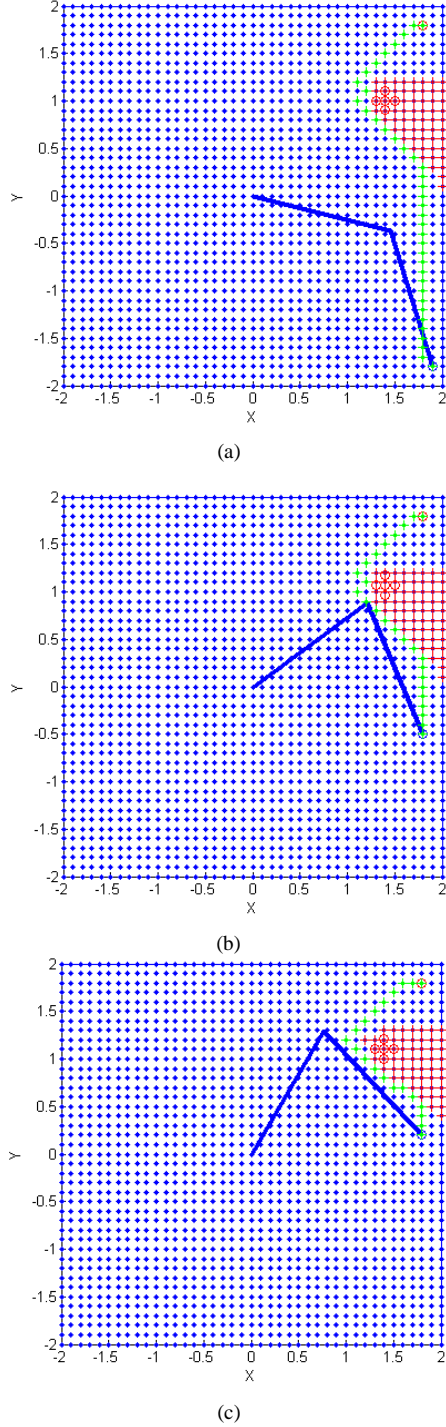


Figure 4. 2-DOF robot manipulator path planning in dynamic environment. The motion sequences are from (a) to (f).

Figure 4 example, demonstrates capability of Grid Roadmap method for real time path planning. Additionally, path smoothing can be applied on resultant path.

## VI. ADVANTAGES OF GRM PATH PLANNING METHOD

Grid Roadmap (GRM) designs an edge less roadmap, accordingly the processing on edge collision checking will confisate. By applying this method instead of DRM, initialization time for edge establishment will eliminate and result a great time save in initialization steps. A brief investigation shows that most path planning methods

desire C-Space. One of the key advantages of GRM real time path planning in dynamic environments is initialization and path search of GRM, which is on workspace with no need for C-Space. Path planning only on workspace eliminates mapping to C-space and vice versa and we save a vast amount of preprocessing time. Proposed method facilitates choosing planner for path planning. Most methods like Focused D\* [12], RRT [13] and RRT based methods [14, 15, 16] are able to achieve desired path on GRM.

We acquire initialization time of GRM and DRM methods on 2-DOF robot arm using MATLAB. Resultant time for DRM is given below:

- PRM generation with 800 random points in C-Space and each point connected to 4 nearest neighbors: 7.081 s.
- Mapping edges and nodes to workspace cells, with 20 points on manipulator elbow and 10\*10 workspace cells: 132.032 s.

Total time for DRM initialization is: 139.113 s.

And resultant time for GRM is:

- “Mapping matrix” creation on a 40\*20 GRM with 150 points on manipulator elbow: 12.61 s

Total time for GRM initialization is: 12.61 s.

And the system properties are:

Pentium IV, 3 GHz CPU and 512 MB of RAM.

We consider nodes and edges with cells are relatively proportion together to have a clear comparison. The achieved resultant times indicates that, the most desired time for DRM method caused by mapping edges and nodes to workspace which is satisfied by GRM method. Accordingly, GRM will be an advance alternate for PRM, DRM and many other PRM based path planning methods. Additionally, Path search for GRM method takes 54 ms and Random Path search for DRM method takes 535 ms. Twice methods path search execution time is suitable for real time path planning in changing environments.

## VII. CONCLUSION

In this paper, we have presented a new Grid Roadmap method for path planning by comparing with DRM and PRM based path planning methods. Proposed roadmap is free of edges and consequently no edge collision checking desired. Hence, improvement in preprocessing time is successfully achieved. We have designed Grid Roadmap on workspace with no desire for configuration space. Consequently, mapping to configuration space and returning to workspace is omitted. Accordingly, we have a great alleviation on initialization time. Additionally, most planners are able to create collision free path on Grid Roadmap.

## REFERENCES

- [1] P. J. McKerrow, Robotics, Addison Wesley, pp. 507-515, 1992.
- [2] M. W. Spong, S. Hutchinson and M. Vidyasagar, Robot Modeling and Control, Wiley, 2005.
- [3] P. Leven and S. Hutchinson, “A Framework for Real-time Path Planning in Changing Environments,” The International Journal of Robotics Research, Vol. 21, No. 12, pp. 999-1030, December 2002.
- [4] T. Kunz, U. Reiser, M. Stilman and A. Verl, “Real-Time Path Planning for a Robot Arm in Changing Environments,” International Conference on Intelligent Robots and Systems, October 2010.
- [5] R. Bohlin and L. E. Kavraki, “Path planning using lazy PRM,” IEEE International Conference on Robotics and Automation. Vol. 1, pp. 521-528, 2000.
- [6] R. Singh, R. M. Voyles, D. Littau and N. P. Papanikolopoulos, “Shape Morphing-Based Control of Robotic Visual Servoing,” Autonomous Robots–AROBOTS, Vol. 10, No. 3, pp. 317-338, 2001.
- [7] Sh. Takahashi, Y. Kokjima and R. Ohbuchi, “Explicit Control of Topological Transitions in Morphing Shapes of 3D Meshes,” Pacific Conference on Computer Graphics and Applications - PG, pp. 70-81, 2001.
- [8] Y. Weng, W. Xu, Y. Wu, K. Zhou and B. Guo, “2D shape deformation using nonlinear least squares optimization, The Visual Computer,” Vol. 22, No. 9-11, pp. 653-660, 2006.
- [9] Y. K. Hwang and N. Ahuja, “A potential field approach to path planning,” IEEE Transaction on Robotics and Automation, Vol. 8, pp. 23-32, February 1992.
- [10] H. Xiaoxi and Ch. Leiting, “Path Planning Based on Grid-Potential Fields,” International Conference on Computer Science and Software Engineering, Vol. 2, pp. 1114-1116, December 2008.
- [11] M. H. Mabrouk, C. R. McInnes, “Wall Following to Escape Local Minima for Swarms of Agents Using Internal States and Emergent Behaviour,” Proceedings of the World Congress on Engineering, Vol. I, July 2008.
- [12] A. Stentz, “The Focused D\* Algorithm for Real-Time Replanning,” In Proceedings of the International Joint Conference on Artificial Intelligence, August 1995.
- [13] J. J. Kuffner and S. M. LaValle, “RRT-Connect: An Efficient Approach to Single-Query Path Planning,” International Conference on Robotics & Automation, April 2000.
- [14] R. Pepy and A. Lambert, “Safe Path Planning in an Uncertain-Configuration Space using RRT,” International Conference on Intelligent Robots and Systems, October 2006.
- [15] B. Burns and O. Brock, “Single-Query Motion Planning with Utility-Guided Random Tree,” IEEE International Conference on Robotics and Automation, April 2007.
- [16] C. Fragkopoulou and A. Graser, “Extended RRT algorithm with dynamic N-dimensional cubic diamonds,” International Conference on Optimization of Electrical and Electronic Equipment, OPTIM, 2010.



# A New Method for Estimating the Maximum Allowable Delay in Networked Control of Bounded Nonlinear Systems

Ashraf F. Khalil<sup>1</sup>, Jihong Wang<sup>2</sup>

<sup>1</sup> School of Electrical, Electronic and Computer Engineering, University of Birmingham, B15 2TT UK

E-mail: afk894@bham.ac.uk

<sup>2</sup> School of Engineering, University of Warwick, Coventry, CV4 7AL UK

E-mail: jihong.wang@warwick.ac.uk

**Abstract.** In Networked Control Systems (NCSs) the information is exchanged through a real-time communication network among control system components. So the network induced time delay and data dropouts are unavoidable in NCSs. The time delay may degrade the performance of control systems and even destabilize the systems if the systems are designed without considering the effects of the time delays properly. Once the structure of a NCS is confirmed, it is essential to identify what the maximum time delay is allowed for maintaining the system stability which, in turn, is also associated with the process of controller design. This paper proposes a new method for estimating the maximum allowable time delay in networked control system with norm bounded nonlinearity. Some studies have been reported in estimation of the maximum time-delay allowed for retaining the system stability. However, most of the reported methods are normally over complicated for practical applications. A new stability method based on using the finite difference approximation for the delay term is proposed in this paper for estimating the maximum time-delay tolerance in networked control system with bounded nonlinearity, which has a simple structure and is easy to apply.

**Keywords-**networked control system; stability; nonlinearity; norm bounded; maximum allowable delay bound

## I. INTRODUCTION

The advances in communication and network technology, and the availability of high speed computers have resulted in an increasing interest in Networked Control Systems (NCSs). This type of control systems can be defined as a control system where the control loop is closed through a real-time communication network [1]. The term "Networked Control Systems" first appeared in Gregory C. Walsh's article in 1998 [2]. A typical organization of an NCS is shown in Fig. 1. In Networked Control Systems, the reference input, plant output and control input are exchanged through a real-time communication network. The main advantages of NCSs are modularity, simplified wiring, low cost, reduced weight, decentralization of control, integrated diagnosis, simple installation, quick and easy for maintenance [3], flexible expandability (easy to add/remove sensors, actuators or controllers with low cost). NCSs are able to easily fuse global information to make intelligent decisions over large physical spaces.

As the control loop is closed through a communication network the time delay and data dropout are unavoidable. This may degrade the performance of NCSs or even

destabilize the system. In general, the control systems with time delays can be classified into time delay independent where the stability is not affected by the time delay and time delay dependent where the time delay affects the stability [4]. Time delay, no doubt, increases complexity in analysis and design of NCSs. Conventional control theories built on a number of standing assumptions including synchronized control and non delayed sensing and actuation must be re-evaluated before they can be applied for NCSs [5].

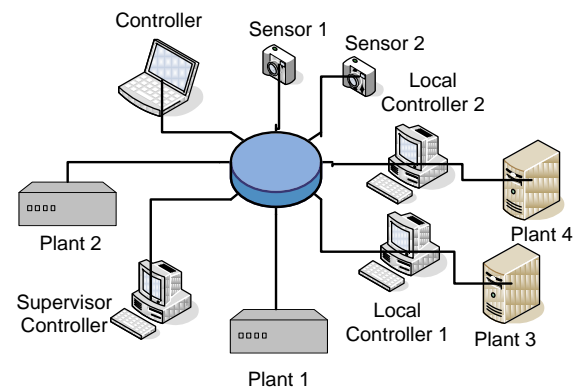


Fig 1 A Typical Networked Control System

In recent years, there are many results for the stability of linear network control systems [6]-[9]. However, there is no much work reported in nonlinear networked control system analysis. Lyapunov functional, Lyapunov-Krasovski functional and Lyapunov-Razumikhin functional based methods are most widely used to study the stability of linear and nonlinear networked control systems where the problem is usually formulated as Linear Matrix Inequalities (LMIs). In [10] Razumikhin and Lyapunov Theorem are used to derive a sufficient stability condition for the stability of a class of nonlinear networked control system. The system under study is the linearized system with a bounded nonlinear function. A discrete-time approach for stabilizing a class of nonlinear system is presented in [11] where the quadratic Lyapunov functional is used to derive a discrete linear controller for the affine nonlinear plant. In [12] a multi-input-multi-output continuous system is studied where the effects of the network induced time delay is modeled as error state-vector which is regarded as a vanishing perturbation. The use of switching Lyapunov functional to derive stability conditions for a networked control system with bounded

nonlinear uncertainty has been studied in [13]. In [14][15] the sampled-data approach is used for a networked control system with nonlinearity and the stability criteria is formulated as LMIs. Their method can be used to calculate the maximum nonlinearity bound for a given time delay and controller but their results are conservative. The maximum nonlinear bound is calculated by solving a constrained convex optimization problem. The Lyapunov-Krasovski functional is used in [16] to derive LMI to study the stability and for designing a stabilizing controller for a networked control system with time-delay, drop-outs and bounded time-varying nonlinearity. The fuzzy-logic approach has been addressed in many papers [17]. The Authors in [17] modeled a class of nonlinear networked control system using Takagi-Sugeno model. They use the approximate model of the discrete nonlinear system to represent the actual system model. In [18] the authors provided new results for stability analysis and stabilization of linear systems with norm bounded nonlinear perturbation. Although the results of the maximum nonlinearity bound are less conservative, the method is limited to free delay systems.

Most of the previously developed approaches require excessive load of computations, and also for higher order systems, the load of computations will increase dramatically. In practice, engineers may find it difficult to apply those available methods in control system design because of the complexity of the methods and lack of guideline in linking between the design parameters and the system performance. Also, the design procedures highly depend on the post-design simulation to determine the design parameters. So there is a demand for a simple design approach with clear guidance for practical applications. The time delay in real-time networks depends strongly on the network protocol and by scheduling the network the time delay can be made smaller and bounded. In this paper a new simple method is proposed for estimating the Maximum Allowable Delay Bound (MADB) in NCS with bounded nonlinearity. The method depends on using the finite difference approximation of the delay term and the problem is formulated as LMI which can be easily solved. Also a simple analytical formula relating the MADB with the maximum nonlinearity is proposed.

The paper starts from description of the proposed method for estimating the maximum time delay for NCS with norm bounded nonlinearity. A few examples are illustrated and the results are compared with that proposed in the previously published literature.

## II. MATHEMATICAL ANALYSIS

A nonlinear system is given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}_p \mathbf{x}(t) + \mathbf{B}_p \mathbf{u}(t) + h(t, \mathbf{x}(t)) \quad (1)$$

where  $\mathbf{x}(t) \in \mathcal{R}^n$  the system state vector and  $\mathbf{u}(t) \in \mathcal{R}^m$  the system control input.  $\mathbf{A}_p \in \mathcal{R}^{n \times n}$  and  $\mathbf{B}_p \in \mathcal{R}^{n \times m}$  are matrices with appropriate sizes.  $h(t, \mathbf{x}(t))$  is the nonlinearity.

The nonlinearity is assumed to be piecewise-continuous function of both  $t$  and  $\mathbf{x}$ .  $h(t, \mathbf{x}(t))$  is uncertain and satisfies the quadratic inequality [14][15];

$$h^T(t, \mathbf{x}(t))h(t, \mathbf{x}(t)) \leq \alpha^2 \mathbf{x}^T(t) \mathbf{H}^T \mathbf{H} \mathbf{x}(t) \quad (2)$$

Where  $\alpha > 0$  is the nonlinearity bounding parameter and  $\mathbf{H}$  is a constant matrix. For any given  $\mathbf{H}$ ;

$$\begin{aligned} \mathbf{H}_\alpha &= \{h: \mathcal{R}^{n+1} \rightarrow \mathcal{R}^n \mid h^T(t, \mathbf{x}(t))h(t, \mathbf{x}(t)) \\ &\leq \alpha^2 \mathbf{x}^T(t) \mathbf{H}^T \mathbf{H} \mathbf{x}(t) \text{ for all } (t, \mathbf{x}) \in \mathcal{R}_+ \times \mathcal{R}^n\} \end{aligned}$$

The constraint (2) can be interpreted as [18];

$$\|h(t, \mathbf{x}(t))\| \leq \alpha \|\mathbf{H} \mathbf{x}(t)\| \quad (3)$$

Stabilizing the system with a linear controller which is given by;

$$\mathbf{u}(t) = \mathbf{K} \mathbf{x}(t - \tau) \quad (4)$$

A typical networked control system model is shown in Fig 2. The time delay may be constant, variable or even random. In NCSs, the time delay is composed of the time delay from sensors to controllers, time delay in controller and controllers to actuators time delay, which is given by:

$$\tau = \tau_{sc} + \tau_c + \tau_{ca} \quad (5)$$

where  $\tau_{sc}$  the time delay between the sensor and controller  $\tau_c$  the time delay in the controller,  $\tau_{ca}$  the time delay from the controller to the actuator. For a general formulation the packet dropouts can be incorporated in (5):

$$\tau = \tau_{sc} + \tau_c + \tau_{ca} + dh \quad (6)$$

where  $d$  is the number of dropouts and  $h$  the sampling period. And by (6) the data dropouts can be considered as a special case of time delay [9]. It is supposed that the following hypotheses hold.

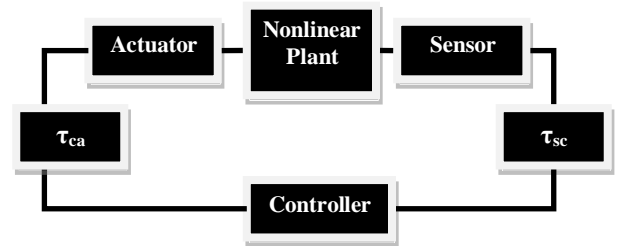


Fig. 2 A Networked Control System Model

### Hypothesis 1 (H.1):

- i. Sensors are clock driven.
- ii. The controllers and actuators are event driven.
- iii. The data are transmitted as a single packet.
- iv. The old packets are discarded.
- v. All the states are available for measurements and hence for transmission.

- vi. The time delay  $\tau$  is small enough to be less than one unit of its measurement.

Applying (4) into (1);

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}_p \mathbf{x}(t) + \mathbf{B}_p \mathbf{K} \mathbf{x}(t - \tau) + h(t, \mathbf{x}(t)) \\ \dot{\mathbf{x}}(t) &= (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \mathbf{x}(t) \\ &\quad + \mathbf{B}_p \mathbf{K} (\mathbf{x}(t - \tau) - \mathbf{x}(t)) + h(t, \mathbf{x}(t))\end{aligned}\quad (7)$$

If H.1 holds then the time delay term can be approximated using the finite difference approximation by Taylor Series expansion. The expression for  $\mathbf{x}(t - \tau)$  can be obtained by Taylor Expansion as:

$$\mathbf{x}(t - \tau) = \sum_{n=0}^{\infty} (-1)^n \frac{\tau^n}{n!} \mathbf{x}^{(n)}(t) \quad (8)$$

Where  $\mathbf{x}^{(n)}(t)$  is the  $n^{th}$  order derivative. The second order approximation of the delay term is given by;

$$\begin{aligned}\mathbf{x}(t - \tau) &= \mathbf{x}(t) - \tau \dot{\mathbf{x}}(t) + (\tau^2 / 2) \ddot{\mathbf{x}}(t) + \mathbf{R}_3(\mathbf{x}, \tau) \\ \mathbf{x}(t - \tau) &\approx \mathbf{x}(t) - \tau \dot{\mathbf{x}}(t) + (\tau^2 / 2) \ddot{\mathbf{x}}(t)\end{aligned}\quad (9)$$

From (9) it can be seen that  $\mathbf{R}_3(\mathbf{x}, \tau)$  depends on the time delay,  $\tau$ , and the higher order derivatives of  $\mathbf{x}(t)$  which can be neglected if the time delay and the norm of  $\mathbf{R}_3(\mathbf{x}, \tau)$  are small. For small time delay and slowly time varying nonlinear perturbation the second derivative can be approximated as:

$$\ddot{\mathbf{x}}(t) \approx (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \dot{\mathbf{x}}(t) \quad (10)$$

Substituting (9) and (10) into (7);

$$\begin{aligned}\dot{\mathbf{x}}(t) &\cong (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \mathbf{x}(t) \\ &\quad + \mathbf{B}_p \mathbf{K} (-\tau \dot{\mathbf{x}}(t) + (\tau^2 / 2) (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \dot{\mathbf{x}}(t)) + h(t, \mathbf{x}(t)) \\ \dot{\mathbf{x}}(t) &\cong [\mathbf{I} + \tau \mathbf{B}_p \mathbf{K} (\mathbf{I} - 0.5\tau (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}))]^{-1} (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \mathbf{x}(t) \\ &\quad + [\mathbf{I} + \tau \mathbf{B}_p \mathbf{K} (\mathbf{I} - 0.5\tau (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}))]^{-1} h(t, \mathbf{x}(t))\end{aligned}\quad (11)$$

Equation (11) can be written as:

$$\dot{\mathbf{x}}(t) \cong \mathbf{N}(\tau) (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \mathbf{x}(t) + g(t, \mathbf{x}(t)) \quad (12)$$

where;

$$\begin{aligned}g(t, \mathbf{x}(t)) &= \mathbf{N}(\tau) h(t, \mathbf{x}(t)) = [\mathbf{I} + \mathbf{M}(\tau)]^{-1} h(t, \mathbf{x}(t)) \\ &= [\mathbf{I} + \tau \mathbf{B}_p \mathbf{K} (\mathbf{I} - 0.5\tau (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}))]^{-1} h(t, \mathbf{x}(t))\end{aligned}$$

According to (2) with the time delay the quadratic inequality can be written as;

$$\begin{aligned}g^T(t, \mathbf{x}(t)) g(t, \mathbf{x}(t)) \\ \leq \alpha^2 \mathbf{x}^T(t) \mathbf{H}^T \mathbf{N}(\tau)^T \mathbf{N}(\tau) \mathbf{H} \mathbf{x}(t)\end{aligned}\quad (13)$$

which can be interpreted as;

$$\|g(t, \mathbf{x}(t))\| \leq \alpha \|\mathbf{N}(\tau) \mathbf{H} \mathbf{x}(t)\| \leq \alpha \|\mathbf{N}(\tau)\| \|\mathbf{H}\| \|\mathbf{x}(t)\| \quad (14)$$

The constraint can be written as;

$$\mathbf{z}(t)^T \begin{bmatrix} -\alpha^2 \mathbf{H}^T \mathbf{N}(\tau)^T \mathbf{N}(\tau) \mathbf{H} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \mathbf{z}(t) \leq 0 \quad (15)$$

$$\text{where; } \mathbf{z}(t) = \begin{bmatrix} \mathbf{x}(t) \\ g(t, \mathbf{x}(t)) \end{bmatrix}$$

Choosing the quadratic Lyapunov functional candidate and taking its derivative;

$$\begin{aligned}\dot{\mathbf{V}}(x) &= \mathbf{x}^T(t) \mathbf{P} \dot{\mathbf{x}}(t) + \dot{\mathbf{x}}^T(t) \mathbf{P} \mathbf{x}(t) \\ &= \mathbf{x}^T \mathbf{P} (\mathbf{N}(\tau) (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \mathbf{x}(t) + g(t, \mathbf{x}(t))) \\ &\quad + (\mathbf{x}(t)^T (\mathbf{A}_p + \mathbf{B}_p \mathbf{K})^T \mathbf{N}(\tau)^T + g^T(t, \mathbf{x}(t))) \mathbf{P} \mathbf{x} \\ \dot{\mathbf{V}}(x) &= \mathbf{x}(t)^T (\mathbf{P} \mathbf{N}(\tau) (\mathbf{A}_p + \mathbf{B}_p \mathbf{K}) \\ &\quad + (\mathbf{A}_p + \mathbf{B}_p \mathbf{K})^T \mathbf{N}(\tau)^T \mathbf{P}) \mathbf{x}(t) \\ &\quad + g^T(t, \mathbf{x}(t)) \mathbf{P} \mathbf{x}(t) + \mathbf{x}(t)^T \mathbf{P} g(t, \mathbf{x}(t))\end{aligned}\quad (16)$$

This can be written as;

$$\mathbf{z}(t)^T \begin{bmatrix} \mathbf{A}^T \mathbf{N}(\tau)^T \mathbf{P} + \mathbf{P} \mathbf{N}(\tau) \mathbf{A} & \mathbf{P} \\ \mathbf{P} & \mathbf{0} \end{bmatrix} \mathbf{z}(t) < 0 \quad (17)$$

Where;  $\mathbf{A} = (\mathbf{A}_p + \mathbf{B}_p \mathbf{K})$

Following the approach in [18] by combining (15) and (17) we get;

$$\begin{bmatrix} \mathbf{A}^T \mathbf{N}(\tau)^T \mathbf{P} + \mathbf{P} \mathbf{N}(\tau) \mathbf{A} + \alpha^2 \mathbf{H}^T \mathbf{N}(\tau)^T \mathbf{N}(\tau) \mathbf{H} & \mathbf{P} \\ \mathbf{P} & -\epsilon \mathbf{I} \end{bmatrix} < 0$$

Letting  $\mathbf{Y} = \epsilon \cdot \mathbf{P}^{-1}$ ,  $\epsilon > 0$ , and using Schur complement we finally get;

$$\begin{bmatrix} \mathbf{N}(\tau) \mathbf{A} \mathbf{Y} + \mathbf{Y} \mathbf{A}^T \mathbf{N}(\tau)^T & \mathbf{I} & \mathbf{Y} \mathbf{H}^T \mathbf{N}(\tau)^T \\ \mathbf{I} & -\mathbf{I} & \mathbf{0} \\ \mathbf{N}(\tau) \mathbf{H} \mathbf{Y} & \mathbf{0} & -\gamma \mathbf{I} \end{bmatrix} < 0 \quad (18)$$

where  $\gamma = 1/\alpha^2$

#### Theorem 1

System (1) and controller (4) with a given time delay is robustly stable with degree  $\alpha$  if the following is feasible

Minimize  $\gamma$

Subject to  $\mathbf{Y} > 0$  and (18)

The optimization problem in Theorem 1 is quasi-convex optimization in  $\gamma$  which can be solved easily using Matlab LMI Toolbox. For systems with small time delays where the first derivative approximation can be used the matrix  $\mathbf{M}(\tau)$  can be approximated as  $\mathbf{M}(\tau) \approx \tau \mathbf{B} \mathbf{K}$ , which can lead to more conservative results.

### Corollary 1

Let H.1 holds, then the nonlinear system (1) with the controller (4) is robustly stable with degree  $\alpha$  if

$$\alpha < \frac{\lambda_{\min}(\mathbf{Q})}{2\|\mathbf{P}\|_2 \cdot \|\mathbf{H}\|_2 \cdot \|\mathbf{N}(\tau)\|}$$

*Proof*

Choosing a Lyapunov functional candidate as:

$$\mathbf{V}(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{x} > 0 \quad \forall \quad \mathbf{x} \neq \mathbf{0} \quad (19)$$

The objective for the next step is to find the range of  $\tau$  that will ensure  $\dot{\mathbf{V}}(\mathbf{x}) < 0$  for  $\forall \quad \mathbf{x} \neq \mathbf{0}$  [19][20]. Taking the derivative of (19) along with the system trajectory (12),

$$\begin{aligned} \dot{\mathbf{V}}(\mathbf{x}) &= \mathbf{x}^T(t) \mathbf{P} \dot{\mathbf{x}}(t) + \dot{\mathbf{x}}^T(t) \mathbf{P} \mathbf{x}(t) \\ &= \mathbf{x}^T \mathbf{P} (\mathbf{N}(\tau)(\mathbf{A} + \mathbf{BK})\mathbf{x}(t) + \mathbf{N}(\tau)h(t, \mathbf{x}(t))) \\ &\quad + (\mathbf{x}(t)^T (\mathbf{A} + \mathbf{BK})^T \mathbf{N}(\tau)^T + h^T(t, \mathbf{x}(t)) \mathbf{N}(\tau)^T) \mathbf{P} \mathbf{x} \\ &= \mathbf{x}^T(t) \mathbf{P} \mathbf{N}(\tau)(\mathbf{A} + \mathbf{BK})\mathbf{x}(t) \\ &\quad + \mathbf{x}(t)^T (\mathbf{A} + \mathbf{BK})^T \mathbf{N}(\tau)^T \mathbf{P} \mathbf{x}(t) + 2\mathbf{x}^T(t) \mathbf{P} \mathbf{N}(\tau)h(t, \mathbf{x}(t)) \\ &= \mathbf{x}^T(t) (\mathbf{P} \mathbf{N}(\tau)(\mathbf{A} + \mathbf{BK}) + (\mathbf{A} + \mathbf{BK})^T \mathbf{N}(\tau)^T \mathbf{P}) \mathbf{x}(t) \\ &\quad + 2\mathbf{x}^T(t) \mathbf{P} \mathbf{N}(\tau)h(t, \mathbf{x}(t)) \end{aligned} \quad (20)$$

If there exists  $\mathbf{P} = \mathbf{P}^T > 0$  and  $\mathbf{Q} = \mathbf{Q}^T > 0$ , satisfying:

$$\mathbf{P} \mathbf{N}(\tau)(\mathbf{A} + \mathbf{BK}) + (\mathbf{A} + \mathbf{BK})^T \mathbf{N}(\tau)^T \mathbf{P} = -\mathbf{Q} \quad (21)$$

Substituting (21) into (20) we get:

$$\dot{\mathbf{V}}(\mathbf{x}(t)) = -\mathbf{x}^T(t) \mathbf{Q} \mathbf{x}(t) + 2\mathbf{x}^T(t) \mathbf{P} \mathbf{N}(\tau)h(t, \mathbf{x}(t)) \quad (22)$$

For any  $\alpha > 0$ , there exist  $r > 0$  such that

$$\|h(t, \mathbf{x}(t))\|_2 < \alpha \|\mathbf{H}\|_2 \cdot \|\mathbf{x}(t)\|_2, \quad \forall \|\mathbf{x}(t)\|_2 < r$$

Then;

$$\begin{aligned} \|\mathbf{N}(\tau)h(t, \mathbf{x}(t))\| &< \|\mathbf{N}(\tau)\| \cdot \|h(t, \mathbf{x}(t))\| \\ &< \alpha \|\mathbf{N}(\tau)\| \cdot \|\mathbf{H}\|_2 \cdot \|\mathbf{x}(t)\|_2 \end{aligned} \quad (23)$$

Also we have [21];

$$\mathbf{x}^T(t) \mathbf{Q} \mathbf{x}(t) \geq \lambda_{\min}(\mathbf{Q}) \|\mathbf{x}(t)\|_2^2 \quad (24)$$

Using (23) and (24) into (22) we finally have;

$$\dot{\mathbf{V}}(\mathbf{x}(t)) < -[\lambda_{\min}(\mathbf{Q}) - 2\alpha \|\mathbf{P}\|_2 \|\mathbf{H}\|_2 \|\mathbf{N}(\tau)\|] \|\mathbf{x}(t)\|_2^2 \quad (25)$$

From (25) it can be found that if

$$\alpha < \frac{\lambda_{\min}(\mathbf{Q})}{2\|\mathbf{P}\|_2 \|\mathbf{H}\|_2 \|\mathbf{N}(\tau)\|}$$

then  $\dot{\mathbf{V}}(\mathbf{x}) < 0$ , the system will be robustly stable with degree  $\alpha$ . We can see from Corollary 1 that the MADB decreases with increasing  $\alpha$ . Setting  $\alpha \approx 0$  and

neglecting the second order term then Corollary 1 reduces to Corollary 1 in [22] as follows;

$$\begin{aligned} \tau &< \frac{1}{\|\mathbf{BK}\|} \left[ 1 - \frac{2\alpha \|\mathbf{P}\|_2 \cdot \|\mathbf{H}\|_2}{\lambda_{\min}(\mathbf{Q})} \right] \\ \tau &< \frac{1}{\|\mathbf{BK}\|} \left[ 1 - 0 \frac{2\|\mathbf{P}\|_2 \cdot \|\mathbf{H}\|_2}{\lambda_{\min}(\mathbf{Q})} \right] \rightarrow \tau < \frac{1}{\|\mathbf{BK}\|} \end{aligned}$$

Increasing  $\alpha$  means moving away from the equilibrium point. We notice as we move away from the equilibrium point the MADB decreases. The boundary of

the domain of attraction is when  $\alpha \approx \frac{\lambda_{\min}(\mathbf{Q})}{2\|\mathbf{P}\|_2 \cdot \|\mathbf{H}\|_2} \rightarrow$

$\tau < (1 - 1/\|\mathbf{BK}\|) \rightarrow 0 < \tau < 0$ , which means the MADB on the boundary is approximately zero.

### III. STABILITY ANALYSIS CASE STUDIES

In general, two approaches are applied to controller design for NCSs. The first design approach is to estimate the maximum allowable delay bound for the system and then the network is scheduled to limit the time delay to be less than the MADB. The second approach is to design the controller while taking the time delay and data dropouts into account. In this paper, the first approach has been adopted. In this section, a number of examples are studied to demonstrate the approach proposed and compared with the previously published cases. In particular, the results derived using the method proposed in this paper has been compared with the results using the LMI method given in [14][15].

#### Example 1

The first example has been studied in [14][15] with the sampled-data approach, the system is given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 1 & 1 \\ 0 & 0.99 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 10 \end{bmatrix} \mathbf{u}(t) + h(t, \mathbf{x}(t))$$

$$\text{with } \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

the controller is chosen in [14][15] to be  $\mathbf{K} = [-0.2999 \quad -0.2989]$ . With  $\alpha \approx 0$  using Theorem 1 the MADB is 0.291 s. For a time delay,  $\tau = 0.2509$  s Using Theorem 1 we have:

$$\alpha_{\max} = 0.1533 \quad \text{with } \mathbf{Y} = \begin{bmatrix} 8.191 & -48.094 \\ -48.094 & 311.1583 \end{bmatrix}$$

Using Corollary 1:

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} 0.7213 & -1.2213 \\ -1.2213 & 2.4487 \end{bmatrix} & \mathbf{Q} &= \mathbf{I} \\ \|\mathbf{P}\|_2 &= 3.0809 & \lambda_{\max}(\mathbf{P}) &= 3.0809 \end{aligned}$$

$$\alpha_{\max} = \frac{\lambda_{\min}(\mathbf{Q})}{2\|\mathbf{P}\|_2 \|\mathbf{H}\|_2 \cdot \left\| [\mathbf{I} + \tau \mathbf{B} \mathbf{K} (\mathbf{I} - 0.5\tau(\mathbf{A} + \mathbf{B} \mathbf{K}))]^{-1} \right\|} = 0.0184$$

The maximum nonlinearity bound given in [14][15] is 0.0013. In [24]  $\alpha_{\max} = 0.1636$  with 0.2509 s time delay. It is clear that the results of Theorem 1 are less conservative than the results of Corollary 1. The trajectory of the system is shown in Fig. 3. For the purpose of the comparison the nonlinear function and the initial conditions for the simulation are given by;

$$h(t, \mathbf{x}(t)) = \begin{bmatrix} \alpha x_1 \sin(x_1) \\ 0 \end{bmatrix} \quad \mathbf{x}(0) = [-15 \quad 10]^T$$

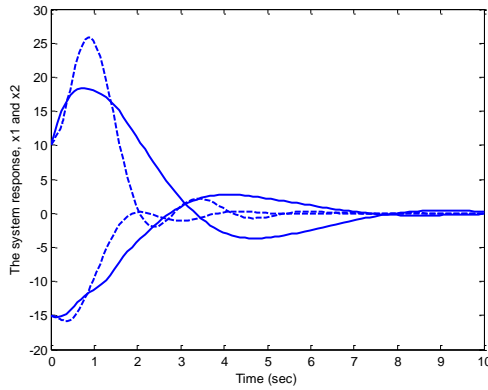


Fig 3 The system response with zero time delay and 0.2509 s time delay

In [14],[15] with 0.22 s time delay and  $\mathbf{K} = [-0.359 \quad -0.317]$ , the maximum nonlinear bound is  $\alpha_{\max} = 0.1365$ , using Corollary 1  $\alpha_{\max} = 0.0252$  while using Theorem 1  $\alpha_{\max} = 0.2555$ . However Corollary 1 and Theorem 1 still give conservative results the method is very easy compared with the method in [14],[15] and [24]. The MADB as a function of the nonlinearity is given in Fig. 4. It can be easily seen that as the nonlinearity increases the MADB decreases.

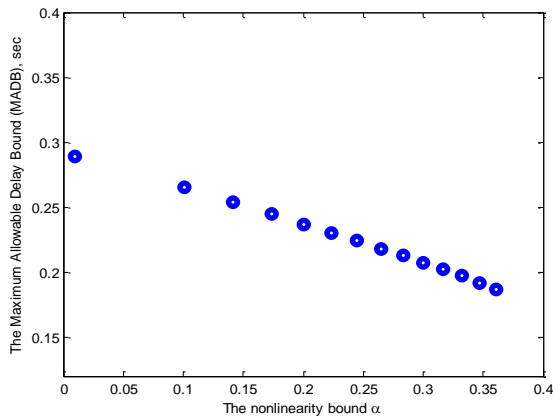


Fig 4 The MADB as a function of the nonlinearity bound using Theorem 1

## Example 2

This example has been studied in [17] with the sampled-data approach, the system is given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -0.5 & 1 \\ 3 & 3 & -2.5 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \\ 0.6 \end{bmatrix} \mathbf{u}(t) + \begin{bmatrix} 0 \\ 0 \\ x_1(t) \sin(x_1(t)) \end{bmatrix}$$

The controller in [17] is designed to be  $\mathbf{K} = [-9.0255 \quad -9.3741 \quad -5.1724]$ . Setting  $\alpha \approx 0$  and using theorem 1 the MADB is 0.0601 s. the maximum nonlinearity bound for the delay free system is 4.3. The MADB as a function of the nonlinearity is shown in Fig. 5. The system response with 0.03 s and 3 nonlinearity is shown in Fig. 6.

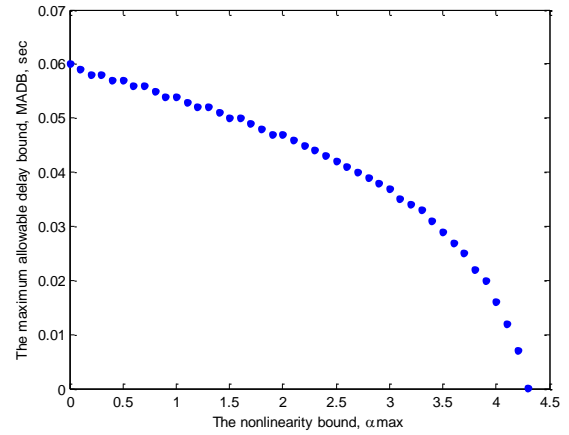


Fig 5 The MADB as a function of the maximum nonlinearity bound using Theorem 1

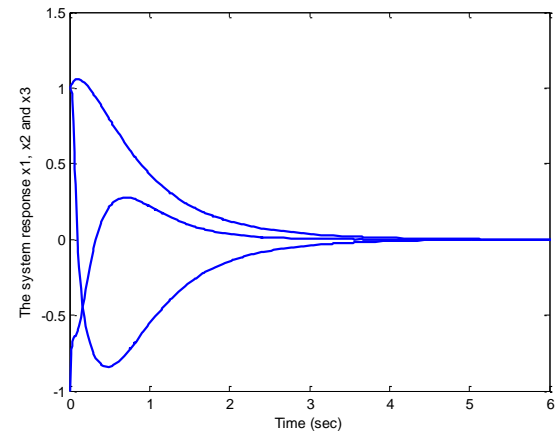


Fig 6 The system response with 0.03 s time delay and  $\alpha = 3$

From Figs. 4 and 6 it is clear that increasing the nonlinearity bound decreases the MADB. The same relation has been noticed in [14][15]. We have carried out many simulation and we found that increasing the

nonlinearity reduces the MADB. Although the method is still conservative but it can be easily applied.

#### IV. CONCLUDING REMARKS

The main contribution of the paper is to have derived a new method for estimating the maximum time delay in NCSs with norm bounded nonlinearity. The most attractive feature of the new method is that it is simple in structure and easy for applications, which can be clearly interpreted to design engineers in industrial sectors. The results obtained in this method are compared with those obtained through the methods introduced in other literatures. The method has demonstrated its merits in using less computation time due to its simple structure and giving less conservative results while showing good agreement with other methods. The method is used to estimate the MADB for a given nonlinearity bound which can be used as a guiding tool for the network scheduling.

#### REFERENCES

- [1] Xiefu Jiang, Qing-Long Han, Shirong Liu, and Anke Xue. "A New  $H_\infty$  Stabilization Criterion for Networked Control Systems", IEEE Transactions On Automatic Control, Vol.53, No.4, 1025-1032, May 2008.
- [2] Gregory C. Walsh and Hong Ye and Linda Bushnell. "Stability Analysis of Networked Control Systems", The Proceedings of the American Control Conference, San Diego, California, USA, Vol. 4, pp. 2876-2880, June 1999.
- [3] Gregory C. Walsh, Hong Ye, and Linda G. Bushnell. "Stability analysis of networked control systems," IEEE Transactions on Control System Technology, Vol. 10, No. 3, 438-446, May 2002.
- [4] Magdi. S. Mahmoud, "Robust Control and Filtering For Time Delay systems", New York: Marcel Dekker, 2000.
- [5] Magdi S. Mahmoud\*, Abdulla Ismail. "Role of Delays in Networked Control Systems", Proceedings of the 10th IEEE International Conference on Electronic Circuits and Systems, Volume 1, 40-43, 2003.
- [6] Dong Yue, Qing-Long Han, and Chen Peng., "State Feedback Controller Design of Networked Control Systems", IEEE Transactions on Circuits and Systems, Vol. 51, No. 11, 640-644, November 2004.
- [7] Bin Tang, Guo-Ping Liu, and Wei-Hua Gui. "Improvement of State Feedback Controller Design for Networked Control Systems", IEEE Transactions on Circuits and Systems, Vol. 55, No. 5, 464-468, May 2008.
- [8] Sun Jian, Liu Guoping, Chen Jie "State Feedback Stabilization of Networked Control Systems", Proceedings of the 27th Chinese Control Conference, Kunming, Yunnan, China, 457-461, July 2008.
- [9] Yuquan Zhang, Qihai Zhong, and Lei Wei, L. "Stability Analysis of Networked Control Systems with Communication Constraints", The Proceedings of Chinese Control and Decision Conference, 335-339, 2008.
- [10] Jun Yang and Xiangdong Wang, "Stability of a Class of Nonlinear Networked Control Systems", The Proceedings of the 5th World Congress on Intelligent Control and Automation, China, 1401-1405, 2004.
- [11] Dan Man, Georgi M. Dimirovski, Jovan D. Stefanovski, and Jun Zhao, "Exponential Stability Synthesis of Networked Nonlinear Control Systems in FMS", 414-419.
- [12] Gregory. C. Walsh, Octavian Beldiman, and Linda G. Bushnell, "Asymptotic Behavior of Nonlinear Networked Control Systems", 1093-1097, IEEE Transactions on Automatic Control, Vol. 46, No. 7,, July 2007.
- [13] Junyan Yu, Long Wang, Mei Yu, Jie Chen and Yingmin Jia, "Robust Controller Design for Networked Control Systems with Nonlinear Uncertainties", 2803-2808, In the Proceedings of the 2009 American Control Conference.
- [14] M. Yu, L. Wang and T. Chu, "Sampled-data Stabilisation of Networked Control Systems with Nonlinearity", 609-614, IEE Proc Control Theory Applications, Vol. 152, No. 6, November 2005.
- [15] M. Yu, L. Wang and T. Chu, "Robust Stabilization of Nonlinear Sampled-data Systems", 3421-3426, In the Proceeding of the 2005 American Control Conference, Portland, USA, 2005.
- [16] Jian Sun and G. P. Liu, "Robust Stabilization of a Class of Nonlinear Networked Control Systems", 2035-2040, In the Proceedings of the 25<sup>th</sup> Chinese Control Conference, 2006.
- [17] Yang Wang, Bin Jiang, and Zehui Mao, "Fault-tolerant Control Design for a kind of Nonlinear Networked Control System with Communication Constraints", 896-901, In the Proceedings of the Chinese Control and Decision Conference, 2009.
- [18] D. D. Siljak and D. M. Stipanovic, "Robust Stabilization of Nonlinear Systems: The LMI Approach", The Journal of Mathematical Problems in Engineering, 461-493, Vol. 6.
- [19] D. P. Goodall, and J. Wang. "Stabilisation of a class of uncertain non-linear affine control systems subject to control constraints", International Journal of Robust and Nonlinear Control, Vol.10, 797-818, 2001.
- [20] J. Wang, J. Pu, P. R. Moore and Z. Zhang. "Modelling study and robust control of air motor systems", International Journal of Control, Vol.71, pp459-476, 1998.
- [21] Hassan K Khalil, Nonlinear Systems, Prentice - Hall International (UK), 1996.
- [22] A.F. Khalil and J.H. Wang, "A New Stability and Time-Delay Tolerance Analysis Approach for Networked Control Systems", The Proceedings of the 49th IEEE Conference on Control and Decision, pp. 4753-4758, 2010.
- [23] Dong-Sung Kim, Young Sam Lee, Wook Hyun Kwon, and Hong Seong Park, "Maximum Allowable Delay Bounds of Networked Control Systems", Journal of Control Engineering Practice, Vol. 11, 1301-1313, 2003.
- [24] Chen Peng, Yu-Chu Tian, and Moses O. Tade, "State Feedback Controller Design of Networked Control Systems with Interval Time-Varying Delay and Nonlinearity", International Journal of Robust and Nonlinear Control, pp. 1-16, 2007.



# Saliency Investigation of PM Brushless AC Motors for High-Frequency Carrier Signal Injection-Based Sensorless Control

Liming Gong and Z. Q. Zhu

Department of Electronic and Electrical Engineering, University of Sheffield  
Sheffield S1 3JD, UK

Elp08lg@sheffield.ac.uk, Z.Q.Zhu@sheffield.ac.uk

**Abstract**—Machine saliency level is a critical concern for high frequency carrier signal injection-based sensorless control of permanent magnet brushless AC (BLAC) motors. To describe the machine saliency level during the whole operation range, machine incremental inductances are required in conventional concepts. To avoid it for machine saliency investigation, a simple experimental method is developed in this paper for practical application, which could easily obtain the machine saliency information without any machine parameters. Based on the measured machine saliency information, Sensorless Safety Operation Area (SSOA) can be obtained, which is in a good agreement with the calculation results. Experimental results confirm the effectiveness of proposed method.

**Keywords**—Permanent magnet brushless AC (BLAC) motors, signal injection, sensorless control, machine saliency.

## I. INTRODUCTION

It is well known that machine saliency is vital for saliency-based sensorless control of permanent magnet brushless AC (BLAC) motors, such as commonly used high frequency (HF) pulsating [1] and rotating carrier signal injection-based methods [2]. In order to describe the machine saliency level, some important terms including anisotropy ratio [3], feasible region [4], and sensorless safety operation area (SSOA) [5] have been introduced. However, the incremental inductances under different load conditions are then required.

According to the definition, incremental self and mutual inductances of permanent magnet BLAC motors can be obtained from finite elemental analysis (FEA) [7]. On the other hand, incremental inductances can also be measured with complex experimental process by injecting HF carrier signal into d- and q-axis, respectively [7].

In order to describe the machine saliency level without requirement of machine parameters, a simple experimental method is proposed in this paper. While locking the rotor at zero position, a pulsating carrier voltage signal is injected in the virtual d-axis ( $d^e$ ), which rotates at fixed speed. From carrier current response, the machine saliency circle, which fully contains machine saliency information, can be easily measured under different load conditions.

With the proposed method, SSOA for the prototype machine can be directly obtained from measured machine saliency circles without the need of any machine parameters. Compared to the calculated SSOA based on machine parameters, the measured SSOA is in good

agreement with calculated results. Therefore, the proposed experimental method provides a simple and effective solution to obtain the machine saliency information.

## II. SALIENCY INFORMATION OF PM BLAC MOTORS

It is well known that a BLAC motor behaves as a pure inductive load when the carrier frequency is high enough than the fundamental excitation frequency. When the cross-saturation effect, which can be represented by the mutual inductances  $L_{dqh}$  and  $L_{qdh}$  in the mathematic model, is considered, the high frequency voltage equations for PM BLAC motors in the synchronous reference frame can be described as [7]:

$$\begin{bmatrix} v_{dh} \\ v_{qh} \end{bmatrix} = \begin{bmatrix} L_{dh} & L_{dqh} \\ L_{qdh} & L_{qh} \end{bmatrix} p \begin{bmatrix} i_{dh} \\ i_{qh} \end{bmatrix} \quad (1)$$

where,  $L_{dh}$  and  $L_{qh}$  are the incremental d- and q-axis self-inductances,  $L_{dqh}$  and  $L_{qdh}$  is the mutual inductances, defined by,

$$\begin{cases} L_{dh} = [\psi_d^r(i_d^r + \Delta i_d^r, i_q^r, \varphi_m) - \psi_d^r(i_d^r, i_q^r, \varphi_m)] / \Delta i_d^r \\ L_{qh} = [\psi_q^r(i_d^r, i_q^r + \Delta i_q^r, \varphi_m) - \psi_q^r(i_d^r, i_q^r, \varphi_m)] / \Delta i_q^r \\ L_{dqh} = [\psi_d^r(i_d^r, i_q^r + \Delta i_q^r, \varphi_m) - \psi_d^r(i_d^r, i_q^r, \varphi_m)] / \Delta i_q^r \\ L_{qdh} = [\psi_q^r(i_d^r + \Delta i_d^r, i_q^r, \varphi_m) - \psi_q^r(i_d^r, i_q^r, \varphi_m)] / \Delta i_d^r \end{cases} \quad (2)$$

For saliency-based sensorless control of PM BLAC motors, machine saliency ( $L_{dh} \neq L_{qh}$ ) resulting from geometric rotor saliency or magnetic saturation is the critical concern. Due to magnetic saturation, the machine saliency varies with fundamental excitation. Based on the definition in (2), the dq-axes incremental self and mutual inductances can be obtained by FEA.

Alternatively, incremental inductances can also be measured with complex experimental process [7]. HF carrier voltage signal is injected into d-axis to obtain the relevant d- and q-axis carrier currents response, and then it is applied to q-axis to record the carrier current response in d- and q-axis, respectively. With the information of injected carrier voltage and measured carrier current response, incremental inductances can be solved from (1). For the prototype machine used in this work, whose parameters are listed in Table I, the measured incremental inductances are shown in Fig.1 [5].

TABLE I PARAMETERS OF INTERIOR PM BLAC MACHINE

Rated voltage (peak)	158V	Rated speed	1000RPM
Rated current (peak)	4.0A	Rated torque	4.0Nm
Rated power	600W	Pole number	6

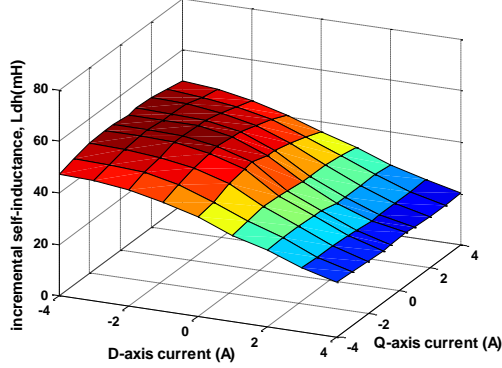
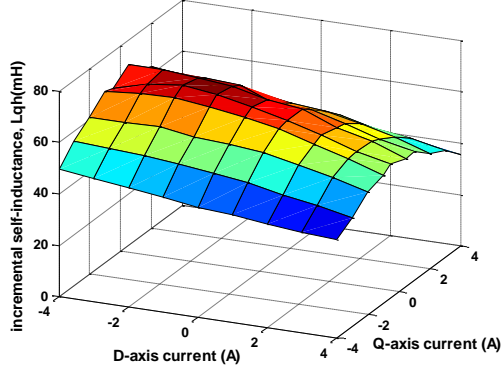
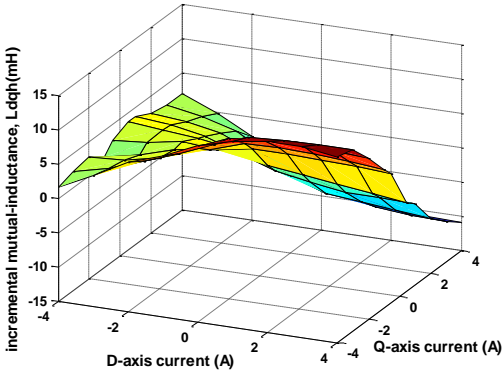
(a)  $L_{dh}$  (mH)(b)  $L_{qh}$  (mH)(c)  $L_{dqh}$  (mH)

Fig.1. Measured incremental self and mutual inductances.

Based on machine incremental inductances, some important terms including anisotropy ratio [3], feasible region [4] and SSOA [5] have been introduced to investigate the machine saliency level. Although the term of anisotropy ratio takes cross-saturation effect into account for saliency investigation, it requires the machine incremental inductances. In [4], the feasible region is used to evaluate the feasibility of sensorless operation, which is defined as the region bounded by the curve of ( $L_{dh}=L_{qh}$ ) in the d-q plane. With consideration of quantization error in AD conversion, SSOA defines a working area in the d-q plane, in which the motor can perform sensorless operation with a guaranteed

performance in steady state [5]. Although SSOA is defined by equivalent negative sequence inductance ( $L_n$ ), the value of  $L_n$  is calculated from incremental inductances as well.

### III. EXPERIMENTAL INVESTIGATION OF MACHINE SALIENCY

Since it is a time-consuming work to obtain machine incremental inductances under different load conditions from FEA or experiment measurement, machine saliency investigation is usually confined to theoretical analysis. To avoid the requirement of machine parameters for machine saliency investigation, a simple experimental method is developed in this paper to investigate machine saliency information for practical applications, which could easily obtain the machine saliency map without any machine parameters.

#### A. Theoretical analysis

In sensorless operation, since the accurate rotor position is unknown, the estimated rotor position is used for control algorithm. Hence, (1) can be transformed to the estimated synchronous reference frame ( $\theta_r^e$ ) from the accurate synchronous reference frame ( $\theta_r$ ) by the transformation matrix,  $T(\Delta\theta)$ , as shown in (3),

$$T(\Delta\theta) = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix} \quad (3)$$

where  $\Delta\theta$  is the estimated position error,  $\Delta\theta = \theta_r - \theta_r^e$ , as shown in Fig.2.

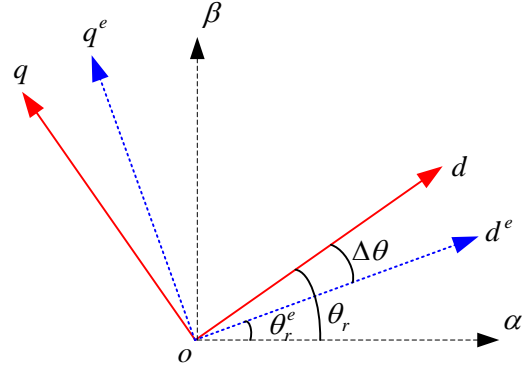


Fig.2. Estimated and accurate synchronous reference frame

For the pulsating carrier signal injection, HF pulsating carrier voltage vector (4) is injected in the estimated d-axis, and then the resultant carrier current in the estimated synchronous reference frame can be derived as (5a):

$$\begin{bmatrix} v_{dh}^e \\ v_{qh}^e \end{bmatrix} = V_c \begin{bmatrix} \cos \alpha \\ 0 \end{bmatrix}, \alpha = \omega_c t + \varphi \quad (4)$$

$$\begin{bmatrix} i_{dh}^e \\ i_{qh}^e \end{bmatrix} = \begin{bmatrix} I_p + I_n \cos(2\Delta\theta + \theta_m) \\ I_n \sin(2\Delta\theta + \theta_m) \end{bmatrix} \cdot \sin \alpha \quad (5a)$$

$$I_p = \frac{V_c}{\omega_c L_p}, I_n = \frac{V_c}{\omega_c L_n} \quad (5b)$$

$$L_p = \frac{L_{dh}L_{qh} - L_{dqh}^2}{L_{sa}}, L_n = \frac{L_{dh}L_{qh} - L_{dqh}^2}{\sqrt{L_{sd}^2 + L_{dqh}^2}} \quad (5c)$$

$$\begin{cases} L_{sa} = (L_{qh} + L_{dh})/2 \\ L_{sd} = (L_{qh} - L_{dh})/2 \end{cases}, \theta_m = \tan^{-1}\left(\frac{-L_{dqh}}{L_{sd}}\right) \quad (5d)$$

where  $I_n$  is a critical parameter for sensorless position estimation, it directly determines the effectiveness of sensorless estimation [5]. For given injected carrier voltage signal ( $V/\omega_c$ ), the value of  $I_n$  is dependent on the machine saliency level ( $L_n$ ), as shown by (5b). The higher the machine saliency level ( $L_n$ ), the higher the value of  $I_n$ ;  $\theta_m$  is the cross-saturation angle induced by the cross-saturation effect, and  $\theta_r$  is the actual rotor position.

### B. Proposed experimental method

Based on theoretical analysis, a simplified experimental method for saliency investigation is developed in this paper. While locking the machine rotor at zero position ( $\theta_r=0^\circ$ ), a pulsating carrier voltage signal (4) with amplitude of 35V and frequency of 330Hz is injected in the virtual d-axis ( $d^e$ ), which rotates at fixed speed (2Hz), as shown in Fig.3. In this case, the carrier current response in the virtual ( $d^e$ - $q^e$ ) reference frame can be expressed as (5a). Meanwhile, a specific fundamental excitation can be applied to validate the saliency variation.

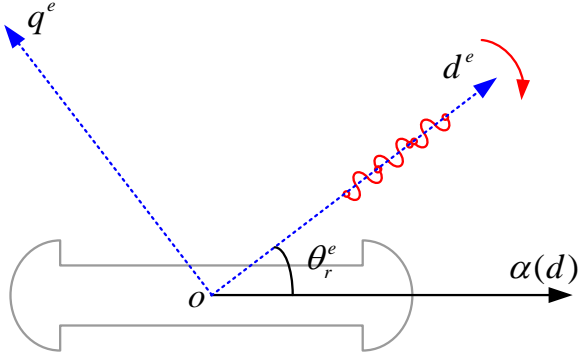


Fig.3. Proposed method for machine saliency investigation.

Without fundamental excitation, the carrier current response in the virtual ( $d^e$ - $q^e$ ) reference frame is measured, as shown in Fig.4. The experimental results are in good agreement with foregoing analysis, which clearly shows that the amplitude of carrier current is modulated by the position difference  $\Delta\theta$  between virtual  $d^e$ -axis and accurate  $d$ -axis.

Utilizing synchronous detection technique [6], the amplitude of  $d^e$ - $q^e$  carrier current can be obtained from (5a), as expressed in (6).

$$\begin{bmatrix} i_{dh}^e \\ i_{qh}^e \end{bmatrix} = LPF\left(\begin{bmatrix} i_{dh}^e \\ i_{qh}^e \end{bmatrix} \cdot 2\sin\alpha\right) = \begin{bmatrix} I_p + I_n \cos(2\Delta\theta + \theta_m) \\ I_n \sin(2\Delta\theta + \theta_m) \end{bmatrix} \quad (6)$$

Fig.5 shows the measured  $d^e$ - $q^e$  carrier current amplitude. From Fig.5(a), when the position difference is  $0^\circ$  or  $180^\circ$ , i.e., the virtual  $d^e$ -axis for injection is aligned with accurate  $d$ -axis,  $d^e$ -axis current amplitude reaches the maximum value, while  $q^e$ -axis current amplitude is close

to zero. From (6), it can be concluded that the cross-saturation angle  $\theta_m$  is nearly zero, i.e., the cross-saturation effect is negligible without fundamental excitation. Furthermore, the amplitude modulation of carrier current undergoes two cycles per single electrical cycle of position, which reveals the angle ambiguity of  $\pi$  in sensorless position estimation. When the  $d^e$ - $q^e$  carrier current amplitude variations are combined together in Fig.5(b), the formed circle is designated as machine saliency circle in this paper, which clearly shows the machine saliency information. Corresponding to equation (6), the center location of machine saliency circle is determined by the value of  $I_p$ , while the radius of saliency circle is dependent on the value of  $I_n$ . Therefore, the scale of the saliency circle indicates the machine saliency level. The longer the radius of saliency circles, the higher the saliency level, and vice versa.

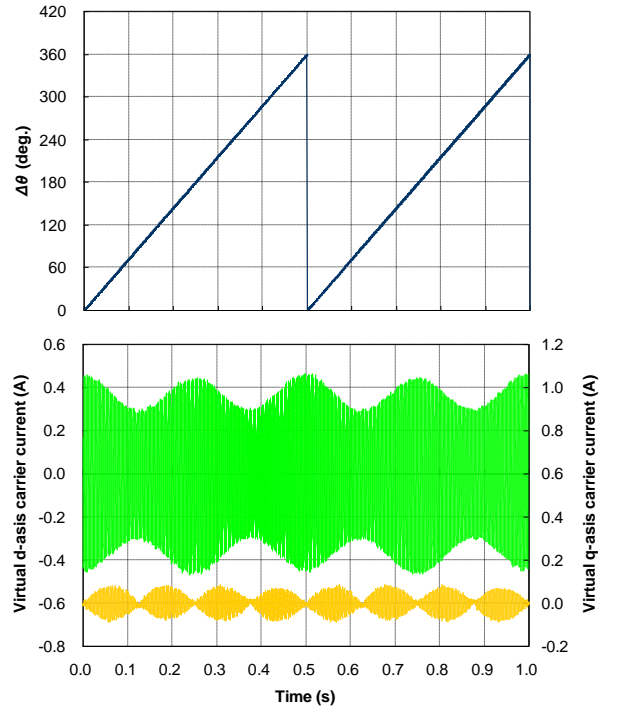
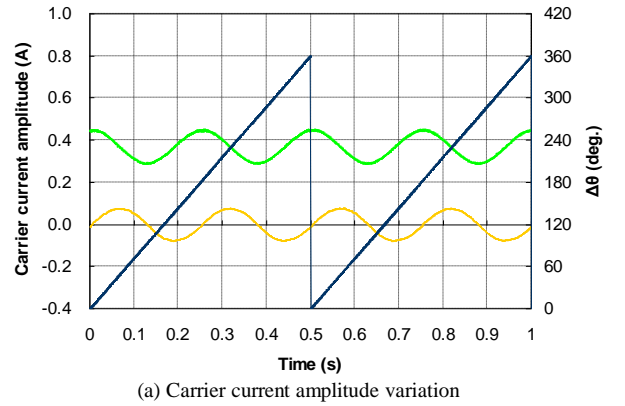


Fig.4 Measured carrier current response in virtual  $d^e$ - $q^e$  reference frame (without fundamental excitation).



(a) Carrier current amplitude variation

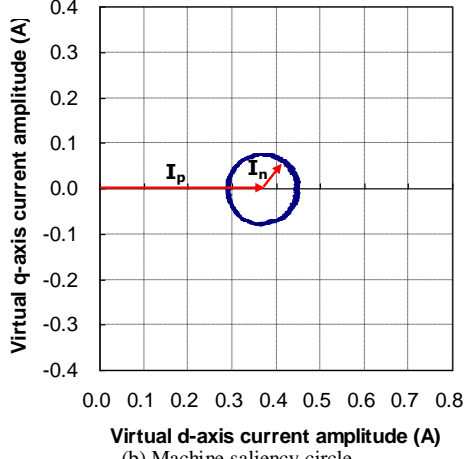


Fig.5. Measured carrier current amplitude variation with position difference (without fundamental excitation).

In the same way, the machine saliency information under different load conditions can be measured, as shown in Fig.6 - Fig.9. From the experimental results, it can be seen that the radius of saliency circle changes significantly with d-axis current, which reveals that d-axis current makes much contributions to the machine saliency level for the prototype machine. Positive d-axis current leads to higher saliency level, and vice versa. Meanwhile, the measured cross-saturation angle shows that cross-saturation effect is mostly dependent on q-axis current, rather than d-axis current.

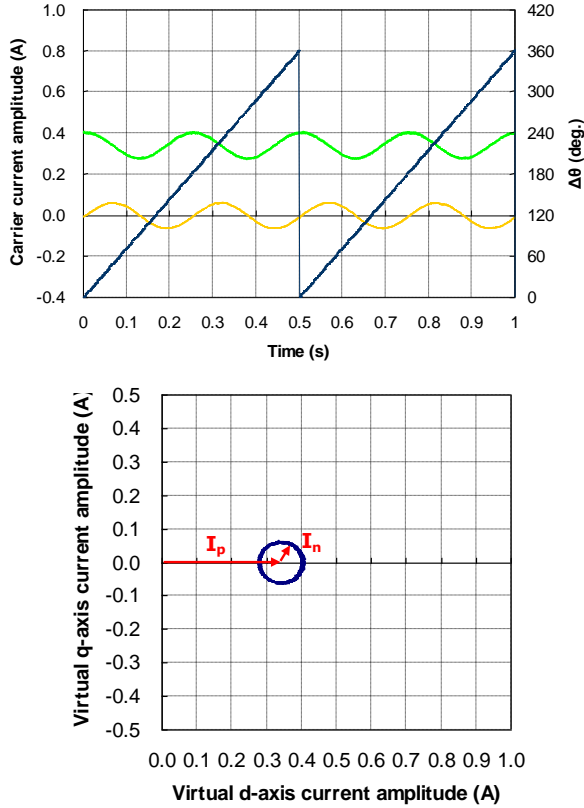


Fig.6. Measured carrier current amplitude variation with position difference ( $i_d = -4A$ ,  $i_q = 0A$ ).

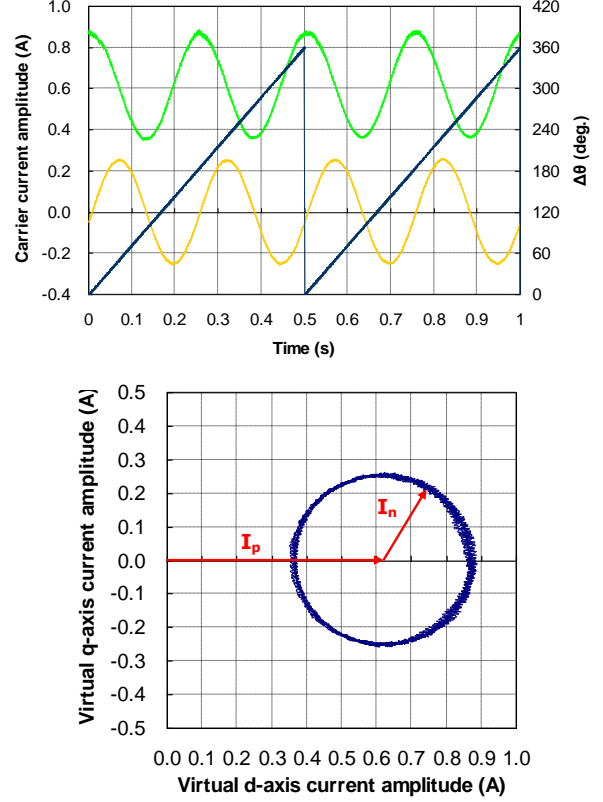


Fig.7. Measured carrier current amplitude variation with position difference ( $i_d = +4A$ ,  $i_q = 0A$ ).

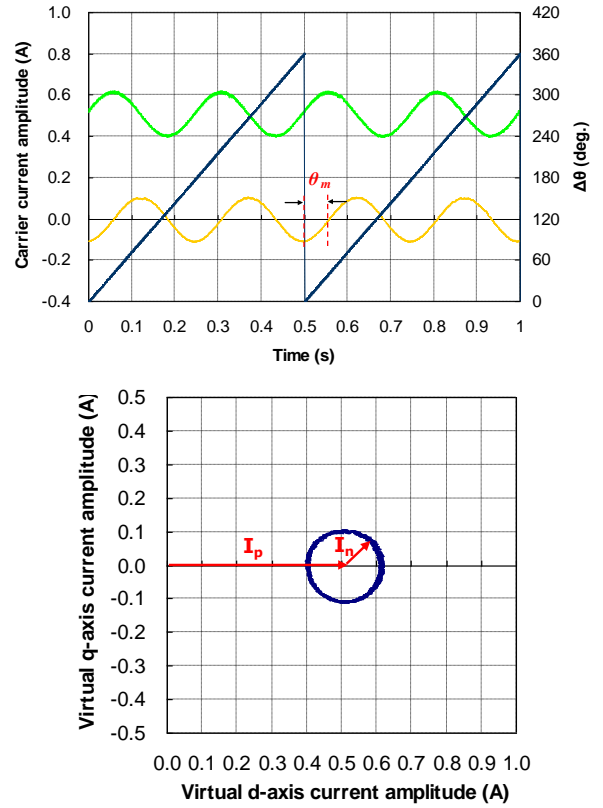


Fig.8. Measured carrier current amplitude variation with position difference ( $i_d = 0A$ ,  $i_q = -4A$ ).



#### IV. CONCLUSION

Machine saliency is critical for HF carrier signal injection-based sensorless control of permanent magnet BLAC motors. To investigate the machine saliency information, some important terms including anisotropy ratio, feasible region and SSOA have been introduced. However, machine incremental inductances are then required, leading to that machine saliency investigation is usually confined to theoretical analysis.

To avoid the requirement of machine parameters, a simple experimental method is developed in this paper to investigate machine saliency information for practical application. With the proposed method, the machine saliency circle, which fully contains machine saliency information, can be easily measured. Based on the measured machine saliency circles, SSOA can be obtained, which is in a good agreement with the calculation results.

As a result, the proposed experimental method provides a simple and effective solution to measure the machine saliency information without machine parameters requirement.

#### REFERENCES

- [1] P.L. Jansen and R.D. Lorenz, "Transducerless position and velocity estimation in induction and salient AC machines," *IEEE Trans. Industry Applications*, vol. 31, no. 2, pp. 240–247, Mar./Apr. 1995.
- [2] M. J. Corley and R. D. Lorenz, "Rotor position and velocity estimation for a salient-pole permanent magnet synchronous machine at standstill and high speed," *IEEE Trans. Industry Applications*, vol. 34, no. 4, pp. 784–789, Jul./Aug. 1998.
- [3] P. Guglielmi, M. Pastorelli, and A. Vagati, "Impact of cross-saturation in sensorless control of transverse-laminated synchronous reluctance motors," *IEEE Trans. Industrial Electronics*, vol. 53, no. 2, pp. 429–439, Apr. 2006.
- [4] N. Bianchi, S. Bolognani, J.H. Jang, and S.K. Sul, "Comparison of PM Motor Structures and Sensorless Control Techniques for Zero-Speed Rotor Position Detection," *IEEE Trans. Power Electronics*, vol. 22, no. 6, pp. 2466–2475, Nov. 2007.
- [5] Z.Q. Zhu, and L.M. Gong, "Investigation of effectiveness of sensorless operation in carrier signal injection-based sensorless control methods," *IEEE Trans. Industrial Electronics*, vol. 58, no. 8, pp. 3431–3439, Aug. 2011.
- [6] A. Madani, J. P. Barbot, F. Colamartino, and C. Marchand, "Reduction of torque pulsations by inductance harmonics identification of a permanent-magnet synchronous machine," *Proceedings of the 4th IEEE Conference on Control Applications*, 1995, pp. 787–792.
- [7] Y. Li, Z.Q. Zhu, D. Howe, C.M. Bingham, "Modeling of cross-coupling magnetic saturation in signal-injection-based sensorless control of permanent-magnet brushless AC motors," *IEEE Trans. Magnetics*, vol. 43, no. 6, pp. 2552–2554, Jun. 2007.



# $H_\infty$ filter design of networked systems with uncertain accessing probabilities and quantisation

Hongbo Song<sup>1</sup>, Li Yu<sup>1</sup> and Guo-ping Liu<sup>2</sup>

<sup>1</sup>Department of Automation, Zhejiang University of Technology, Hangzhou, China

<sup>2</sup>Faculty of Advanced Technology, University of Glamorgan, Pontypridd, UK

Email, di7gan\_shb@hotmail.com, lyu@zjut.edu.cn, gpliu@glam.ac.uk

**Abstract**—This paper is concerned with the  $H_\infty$  filtering problem for networked systems with limited accessing measurement and quantisation effects. The plant has multiple quantised measurement outputs and only one of them can be received by the filter at each transmission instant. The considered situation is stochastic accessing with uncertain probabilities based on the observation that estimated probabilities with error bounds instead of exact ones can be obtained in practical systems. The accessing process is modeled as a Bernoulli sequence and the corresponding filtering error system is modeled as a stochastic system with uncertainties both in the system matrix and probabilities. Sufficient conditions are presented for the filtering error system to be stochastically stable with an  $H_\infty$  performance and a filter design method is also presented in terms of liner matrix inequalities. Finally, an illustrative example is given to show the effectiveness of the proposed method.

**Keywords**- networked systems; limited accessing measurement; uncertain probabilities;  $H_\infty$  filtering; quantisation

## I. INTRODUCTION

Networked systems have attracted increasing attentions in the last decade due to its wide applications in engineering fields, such as remote sensing, unmanned aerial vehicles and automobile [1-5]. The use of a network introduces many issues that degrade the performance of a networked system, for example, delay, packet dropout, medium access constraint, quantisation, and so on [2-7]. These issues have been extensively studied in the literature from various aspects, such as stability analysis, controller synthesis and filtering.

In networked systems, plant outputs need to be quantised before transmitted [6]. Furthermore, only limited measurement outputs can be received by the controller/filter due to the medium access constraint [7]. Depend on the employed medium access control protocols, the accessing process can be grouped into determination and stochastic cases. The determination case has been studied in [5,7,11] and the references therein. However, few results are available on the stochastic case, which is also research field deserves investigation. Ethernet is the main motivation for studying the stochastic accessing problem [12-13]. In [12], the stability analysis problem was studied for general nonlinear networked systems with stochastic protocols. In [13], a stochastic sensor selection method was proposed for estimating a signal by minimizing the upper bound on the error variance. As well

known, the  $H_\infty$  scheme is often employed when the statistic of the external noise is not exactly known. The  $H_\infty$  filtering problem for networked system from various aspects were studied in [8-10,18] and the references therein.

On the other hand, in practical systems only estimated probabilities with uncertainty bounds are available instead of exact values for the probabilities. However, in the existing results concerning the  $H_\infty$  filtering problem for networked systems, the probabilities are assumed to be exact known. Thus, we consider the problem of  $H_\infty$  filter design for networked system with uncertain accessing probabilities and quantisation, which is an interesting problem deserves investigation

In this paper, the  $H_\infty$  filtering problem is studied for networked systems with uncertain accessing probabilities and quantisation. The plant has multiple measurement outputs. After quantised, only one of them can be received by the filter at each transmission instant. The accessing process is governed by a Bernoulli sequence with uncertain probabilities. It is assumed that the probabilities are obtained as estimated values with uncertain bounds, which called the element-wise way. A transformation is presented for element-wise and polotopic formulations of the uncertain probabilities. Then the filtering error system is modeled as a stochastic system with uncertainties both in the system matrices and probabilities. The filtering error system is a special case of Markovian Jump System. Uncertain probabilities issues were considered in Markov jumping systems, see for example [15-17], so that some existing results are available to use. Sufficient conditions are presented for the filtering error systems to be stochastically stable with an  $H_\infty$  performance. Based on the conditions, a design method is also presented for the optimal  $H_\infty$  filter. An illustrative example is given to show the effectiveness of the proposed method.

## II. PROBLEM FORMULATION

The structure of the considered networked systems is shown in Fig. 1, where the plant is described by the following linear discrete-time state space model:

$$\begin{cases} x(k+1) = Ax(k) + Bw(k) \\ y(k) = Cx(k) + Dw(k) \\ z(k) = Lx(k) \end{cases} \quad (1)$$

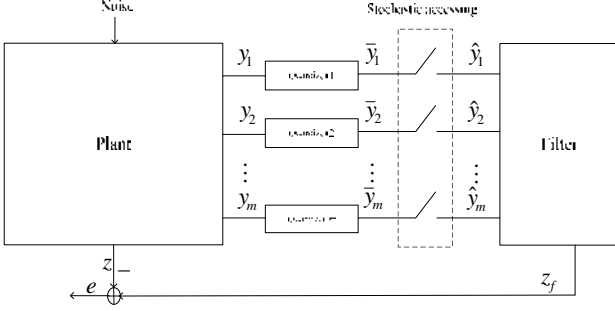


Figure 1. Structure of the considered networked system

where  $x \in R^{n_x}$  is the state of the plant,  $y \in R^m$  is the measured output of the plant,  $\bar{y} \in R^m$  is the output of the quantisers,  $w \in R^{n_w}$  is the external finite-energy disturbance input, and  $z \in R^{n_z}$  is the signal to be estimated.  $A$ ,  $B$ ,  $C$ ,  $D$  and  $L$  are matrices with appropriate dimensions, and  $A$  is assumed to be stable.

The objective is to estimate the signal  $z$  via a filter which is connected to the plant through a network. The filter is a full order one of the following form

$$\begin{cases} x_f(k+1) = A_f x_f(k) + B_f \hat{y}(k) \\ z_f(k) = C_f x_f(k) \end{cases} \quad (2)$$

where  $x_f \in R^{n_x}$  is the state of the filter,  $\hat{y} \in R^m$  is the input of the filter, and  $z_f \in R^{n_z}$  is the estimated signal.  $A_f$ ,  $B_f$  and  $C_f$  are filter parameter matrices to be determined. Let  $y = [y_1 \ y_2 \ \dots \ y_m]^T$ ,  $\bar{y} = [\bar{y}_1 \ \bar{y}_2 \ \dots \ \bar{y}_m]^T$  and  $\hat{y} = [\hat{y}_1 \ \hat{y}_2 \ \dots \ \hat{y}_m]^T$ , and  $y_i$ ,  $\bar{y}_i$  and  $\hat{y}_i$ ,  $\forall i \in M = \{1, 2, \dots, m\}$  are scalars.

We focus on the limited accessing measurement and quantisation issues in this paper. As well known, plant measurement outputs have to be quantised before transmitted in networked systems. The commonly-used logarithmic quantisers are adopted here. The multiple density quantisers are denoted by

$$q(\bullet) = [q_1(\bullet) \ q_2(\bullet) \ \dots \ q_m(\bullet)]^T$$

where  $q_j$ , are called quantiser  $j$ ,  $\forall j \in M$ . The set of the quantisation level of quantiser  $j$  is represented by  $U_j = \{\pm u_i^{(j)} : i = 0, \pm 1, \pm 2, \dots\} \cup \{0\}$ . A quantiser is called logarithmic if  $U_j = \{\pm u_i^{(j)} : u_i^{(j)} = \theta_j^i u_0^{(j)}, i = \pm 1, \pm 2, \dots\} \cup \{u_0^{(j)}\} \cup \{0\}$ ,  $0 < \theta_j < 1$ . and the corresponding  $q_j$  is defined as follows:

$$q_j(v) = \begin{cases} u_i^{(j)}, & \text{if } \frac{1}{1+\xi_j} u_i^{(j)} < v \leq \frac{1}{1-\xi_j} u_i^{(j)}, v > 0 \\ 0, & \text{if } v = 0 \\ -q_j(-v), & \text{if } v < 0 \end{cases} \quad (3)$$

where  $\xi_j = \frac{1-\theta_j}{1+\theta_j}$ .  $\theta_j$  is called the quantisation density of quantiser  $j$  and a smaller  $\theta_j$  means a coarser  $q_j$  [6]. Then we have  $\bar{y} = q(y) = [q(y_1) \ q(y_2) \ \dots \ q(y_m)]^T$ . According to [6],  $\bar{y}$  can be expressed as

$$\bar{y}(k) = (I + H(k))y(k) \quad (4)$$

where

$$H(k) = \text{diag}\{h_1(k), h_2(k), \dots, h_m(k)\}, \quad h_i(k) \in [-\xi_i, \xi_i], \quad \forall i \in M, \quad \text{diag}\{\bullet\} \text{ denotes a block diagonal matrix and } I \text{ is a identity matrix with compatible dimension.}$$

With regard to the limited accessing measurement, we consider the most strict case in which only one element of  $\bar{y}(k)$  can be received by the filter at each time instant without loss of generality. The rest of the elements in  $\bar{y}(k)$  are simply discarded. On the other hand, the elements in  $\hat{y}(k)$  that are not updated are simply set to zero. Then it can be readily seen that if  $\bar{y}_i(k)$  is received by the filter at instant  $k$ , then  $\hat{y}_i(k) = \bar{y}_i(k)$  and  $\hat{y}_j(k) = 0$ ,  $\forall j \in M \setminus \{i\}$ .

Denote by  $\rho(k)$  the subscript index of the element in  $\bar{y}(k)$  which gets access to the network and  $\rho(k)$  takes values in  $M$ . Denote by  $S_i$  the event that  $\bar{y}_i(k)$  is received by the filter. Introduce

$$\Lambda_i = \text{diag}\{\delta(i-1), \delta(i-2), \dots, \delta(i-m)\}, \quad \forall i \in M$$

where  $\delta(l) = \begin{cases} 0, & l \neq 0 \\ 1, & l = 0 \end{cases}$ , then it can be obtained that

$$\hat{y}(k) = \Lambda_{\rho(k)} \bar{y}(k) \quad (5)$$

By (4) and (5) it can be readily obtained that

$$\hat{y}(k) = \Lambda_{\rho(k)} (I + H(k))y(k) \quad (6)$$

As mentioned before, one feature of this paper is that we consider the stochastic case with uncertain probabilities. We model the stochastic accessing process  $\rho(k)$  as a Bernoulli sequence, which is a special case of Markovian chain [4]. It is memoryless and

$$\text{Prob}\{\rho(k+1) = i | \rho(k) = j\} = \text{Prob}\{\rho(k+1) = i\}$$

where  $\text{Prob}\{\bullet\}$  denotes the probability of occurrence of a event. Let  $\phi_i(k) = \text{Prob}\{\rho(k) = i\}$ . In practical systems, the exact values for probabilities are difficult to obtain. However, the estimated values and a bound for the uncertainty can be determined. This is called element-wise description [16]. We denote these by  $\phi_i(k) = \sigma_i + \Delta\sigma_i$ ,  $\Delta\sigma_i \in [-\mu, \mu]$ ,  $\forall i \in M$ , where  $\sigma_i$  and  $\Delta\sigma_i$  are the nominal part and the uncertain part,

respectively, and they should satisfy  $\sum_{i=1}^m \sigma_i = 1$  and  $\sum_{i=1}^m \Delta \sigma_i = 0$ .

**Remark 1:** It should be pointed out that  $\Delta \sigma_i \in [-\mu, \mu]$  rather than  $\Delta \sigma_i \in [-\mu_i, \mu_i]$ . The reason is that if  $\mu_i \neq \mu_j$ ,  $i \in M, j \in M \setminus \{i\}$ , then there exist values in the uncertain parts that do not satisfy  $\sum_{i=1}^m \Delta \sigma_i = 0$ .

Denote  $\eta^T(k) = [x^T(k) \ x_f^T(k)]$  and  $e(k) = z_f(k) - z(k)$ , we obtain by (1), (2) and (6) the following filtering error system:

$$\begin{cases} \eta(k+1) = \tilde{A}_{\rho(k)} \eta(k) + \tilde{B}_{\rho(k)} w(k) \\ e(k) = \tilde{C} \eta(k) \end{cases} \quad (7)$$

where  $\tilde{A}_{\rho(k)} = \begin{bmatrix} A & 0 \\ B_f \Lambda_{\rho(k)} (I + H(k)) C & A_f \end{bmatrix}$ ,

$\tilde{B}_{\rho(k)} = \begin{bmatrix} B \\ B_f \Lambda_{\rho(k)} (I + H(k)) D \end{bmatrix}$  and  $\tilde{C} = [-L \ C_f]$ .

In (7), it can be seen that the probability uncertainties and system uncertainties are coupled. This will make the  $H_\infty$  filter design complex if we use the element-wise description. Fortunately, the element-wise and polytopic descriptions can be equivalently represented and the polytopic description will lead to a simpler design method [17]. An equivalent polytopic description is given as follows.

Denote the vertices by  $V_{ij}$ ,  $\forall i \in M, j \in M \setminus \{i\}$ . Denote the probabilities of occurrence of  $S_s$ ,  $\forall s \in M$  at  $V_{ij}$  by  $[\varepsilon_{ij}^{(1)} \ \varepsilon_{ij}^{(2)} \ \dots \ \varepsilon_{ij}^{(m)}]$ . By the formulation of the element-wise description, we define that at  $V_{ij}$ ,  $\varepsilon_{ij}^{(i)} = \sigma_i - \mu$ ,  $\varepsilon_{ij}^{(j)} = \sigma_j + \mu$ ,  $\varepsilon_{ij}^{(s)} = \sigma_s$ ,  $\forall s \in M \setminus \{i, j\}$ . For example with  $m=3$ , it can be obtained that the probabilities are  $[\sigma_1 \ \sigma_2 - \mu \ \sigma_3 + \mu]$  at  $V_{23}$ .

**Remark 2:** In [16], it is pointed out that the number of vertices of the polytopic description will be extremely large when the mode of Markovian jump system is larger than three. However, in our case, the number of vertices is  $m \times (m-1)$ , which is acceptable to use the polytopic description.

The problem of  $H_\infty$  filter design for the considered networked system is now formulated the  $H_\infty$  performance analysis and filter design problem for the filtering error system (7), which has uncertainties both in the system matrices and stochastic parameters. The main results will be given in the next section.

### III. MAIN RESULTS

In this section, a design method is presented for the desired  $H_\infty$  filter. Before proceeding further, the

following Definition and Lemmas are given as follows first.

**Definition 1** [4]: The filtering error system (7) is said to be stochastically stable with a prescribed  $H_\infty$  performance  $\gamma$  if the following hold:

1) the filtering error system (7) with  $w(k)=0$  is stochastically stable, that is, for any initial condition  $\eta(0)$ ,  $E\{\sum_{i=0}^{+\infty} \|\eta(i)\|^2 | \eta(0)\} < +\infty$  holds, where  $E\{\bullet\}$  is the expectation operator and  $\|\eta(k)\| = \sqrt{\eta^T(k) \eta(k)}$ .

2) under zero initial condition,  $E\{\|e\|_2\} < \gamma \|w\|_2$  where  $\|e\|_2 = \sqrt{\sum_{k=0}^{+\infty} \|e(k)\|^2}$ .

**Lemma 1** [4]: Assume a Markovian jump system is weakly controllable and the Markov parameter is a Bernoulli sequence with  $\alpha_i$  being the probabilities of each mode, then the system is stochastically stable with an  $H_\infty$  performance  $\gamma$  if and only if there exists a matrix  $P > 0$  such that the following inequality holds

$$\sum_{i=1}^m \alpha_i \begin{bmatrix} A_i & B_i \\ C_i & 0 \end{bmatrix}^T \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A_i & B_i \\ C_i & 0 \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & \gamma^2 I \end{bmatrix} < 0 \quad (8)$$

**Remark 3:** Lemma 1 is a special case of the well-known Bounded Real Lemma for Markovian jump system. Note that the weak controllability is only for the necessity [14].

**Lemma 2** [8]: For given appropriate matrices  $\Psi_1$ ,  $\Psi_2$ , and  $\Psi_3$  with  $\Psi_1^T = \Psi_1$ , then  $\Psi_1 + \Psi_3 \Delta(k) \Psi_2 + \Psi_2^T \Delta^T(k) \Psi_3^T < 0$  holds for all  $\Delta^T(k) \Delta(k) \leq I$  if and only if there exists a scalar  $\lambda > 0$  such that  $\Sigma_1 + \lambda \Psi_3^T \Psi_3 + \lambda^{-1} \Psi_2^T \Psi_2 < 0$ .

**Theorem 1:** For given scalars  $\gamma$  and  $\sqrt{\varepsilon_{ij}^{(s)}}$ ,  $\forall i, s \in M$ ,  $j \in M \setminus \{i\}$ , if there exist matrices  $P > 0$ ,  $A_f$ ,  $B_f$  and  $C_f$ , and a scalar  $\lambda$  such that the following inequalities

$$\begin{bmatrix} \Phi_{11} & * & * & * \\ \Phi_{21} & \Phi_{22} & * & * \\ \Phi_{31} & \Phi_{32} & -\Phi_{33} & * \\ 0 & 0 & \Phi_{43} & -\lambda I \end{bmatrix} < 0 \quad (9)$$

hold for all  $\forall i \in M$ ,  $j \in M \setminus \{i\}$ , then the filtering error system (7) is stochastically stable with an  $H_\infty$  performance  $\gamma$ , where

$$\begin{aligned} \Phi_{11} &= -P + \tilde{C}^T \tilde{C} + \lambda \tilde{C}^T \bar{C}, \quad \Phi_{21} = \lambda \bar{C}^T D, \\ \Phi_{22} &= -\gamma^2 I + \lambda D^T D, \quad \bar{C} = [C \ 0], \\ \Phi_{31} &= \Phi_{33} Q \Gamma_1, \quad \Phi_{32} = \Phi_{33} Q \Gamma_2, \quad \Phi_{43} = \Gamma_3 Q \Phi_{33}, \\ \Phi_{33} &= \text{diag}\{P, P, \dots, P\}, \end{aligned}$$

$$Q = \text{diag} \left\{ \sqrt{\varepsilon_{ij}^{(1)}} I \quad \sqrt{\varepsilon_{ij}^{(2)}} I \quad \dots \quad \sqrt{\varepsilon_{ij}^{(m)}} I \right\},$$

$$\Gamma_a = \begin{bmatrix} \Gamma_{a1}^T & \Gamma_{a2}^T & \dots & \Gamma_{am}^T \end{bmatrix}^T, \quad a=1,2,$$

$$\Gamma_3 = \begin{bmatrix} \Gamma_{31}^T & \Gamma_{32}^T & \dots & \Gamma_{3m}^T \end{bmatrix}^T,$$

$$\Gamma_{1s} = \begin{bmatrix} A & 0 \\ B_f \Lambda_s C & A_f \end{bmatrix}, \quad \Gamma_{2s} = \begin{bmatrix} B \\ B_f \Lambda_s D \end{bmatrix}, \quad \Gamma_{3s} = \begin{bmatrix} 0 \\ B_f \Lambda_s \Pi \end{bmatrix},$$

$$\forall s \in M, \quad \Pi = \text{diag} \{ \xi_1, \xi_2, \dots, \xi_m \}.$$

**Proof:** In (8) it can be seen that the Bounded Real Lemma is affine in the probabilities. Since the probabilities are in a convex set, it can be seen that if the inequalities

$$\sum_{s=1}^m \varepsilon_{ij}^{(s)} \begin{bmatrix} \tilde{A}_s & \tilde{B}_s \\ \tilde{C} & 0 \end{bmatrix}^T \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{A}_s & \tilde{B}_s \\ \tilde{C} & 0 \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & \gamma^2 I \end{bmatrix} < 0 \quad (10)$$

hold for all  $i \in M$ ,  $j \in M \setminus \{i\}$ , then by Lemma 1 the filtering error system (7) is stochastically stable with an  $H_\infty$  performance  $\gamma$ .

Since  $\tilde{C}$  is independent on  $\rho(k)$ , so (10) can be rewritten as follows

$$\sum_{s=1}^m \varepsilon_{ij}^{(s)} \begin{bmatrix} \tilde{A}_s & \tilde{B}_s \end{bmatrix}^T P \begin{bmatrix} \tilde{A}_s & \tilde{B}_s \end{bmatrix} - \begin{bmatrix} P - \tilde{C}^T \tilde{C} & 0 \\ 0 & \gamma^2 I \end{bmatrix} < 0 \quad (11)$$

Then by Schur complement we have

$$\Omega_1 + \Omega_3 \Delta(k) \Omega_2 + \Omega_2^T \Delta^T(k) \Omega_3^T < 0 \quad (12)$$

where

$$\Omega_1 = \begin{bmatrix} -P + \tilde{C}^T \tilde{C} & * & * & * & * & * \\ 0 & -\gamma^2 I & * & * & * & * \\ \sqrt{\varepsilon_{ij}^{(1)}} \Gamma_{11} & \sqrt{\varepsilon_{ij}^{(1)}} \Gamma_{21} & -P^{-1} & * & * & * \\ \sqrt{\varepsilon_{ij}^{(2)}} \Gamma_{12} & \sqrt{\varepsilon_{ij}^{(2)}} \Gamma_{22} & 0 & -P^{-1} & * & * \\ \vdots & \vdots & \vdots & \vdots & \ddots & * \\ \sqrt{\varepsilon_{ij}^{(m)}} \Gamma_{1m} & \sqrt{\varepsilon_{ij}^{(m)}} \Gamma_{2m} & 0 & 0 & \dots & -P^{-1} \end{bmatrix},$$

$$\Omega_2 = \begin{bmatrix} 0 & 0 & \sqrt{\varepsilon_{ij}^{(1)}} \Gamma_{31}^T & \sqrt{\varepsilon_{ij}^{(2)}} \Gamma_{32}^T & \dots & \sqrt{\varepsilon_{ij}^{(m)}} \Gamma_{3m}^T \end{bmatrix}^T,$$

$$\Omega_3 = \begin{bmatrix} \tilde{C} & D & 0 & 0 & \dots & 0 \end{bmatrix}, \quad \Delta^T(k) \Delta(k) \leq I.$$

By Lemma 2, Schur complement and some matrix manipulations it can be readily obtained that (9) and (12) are equivalent. Therefore, if (9) hold for all  $\forall i \in M$ ,  $j \in M \setminus \{i\}$ , then the filtering error system (7) is stochastically stable with an  $H_\infty$  performance  $\gamma$ . The proof is completed.

Theorem 1 gives sufficient conditions for the stochastic stability with an  $H_\infty$  performance of the filtering error system (7), however, filter parameters are not available because they are coupled with Lyapunov matrix  $P$ . A design method is presented as follows.

**Theorem 2:** For given scalars  $\gamma$ , and  $\sqrt{\varepsilon_{ij}^{(s)}}$ ,  $i, s \in M$ ,  $j \in M \setminus \{i\}$ , if there exist matrices  $\bar{P}_1, \bar{P}_2, \bar{P}_3, X_1, X_2, X_3, \bar{A}_f, \bar{B}_f$  and  $\bar{C}_f$ , and a scalar  $\lambda$  such that the following inequalities

$$\begin{bmatrix} \Xi_{11} & * & * & * & * \\ \Xi_{21} & \Xi_{22} & * & * & * \\ \Xi_{31} & \Xi_{32} & \Xi_{33} & * & * \\ 0 & 0 & \Xi_{43} & -\lambda I & * \\ \Xi_{51} & 0 & 0 & 0 & -I \end{bmatrix} < 0 \quad (13)$$

hold for all  $\forall i \in M$ ,  $j \in M \setminus \{i\}$ , then the filtering error system (7) is stochastically stable with an  $H_\infty$  performance  $\gamma$ , where

$$\Xi_{11} = \begin{bmatrix} -\bar{P}_1 + \lambda C^T C & * \\ -\bar{P}_2 & -\bar{P}_3 \end{bmatrix}, \quad \Xi_{21} = \begin{bmatrix} \lambda C^T D & 0 \end{bmatrix},$$

$$\Xi_{22} = -\gamma^2 I + \lambda D^T D,$$

$$\Xi_{3a} = Q \begin{bmatrix} \Xi_{3a1}^T & \Xi_{3a2}^T & \dots & \Xi_{3am}^T \end{bmatrix}^T, \quad a=1,2$$

$$\Xi_{43} = Q \begin{bmatrix} \Xi_{431}^T & \Xi_{432}^T & \dots & \Xi_{43m}^T \end{bmatrix}^T,$$

$$\Xi_{31i} = \begin{bmatrix} X_1^T A + \bar{B}_f \Lambda_i C & \bar{A}_f \\ X_3^T A + \bar{B}_f \Lambda_i C & \bar{A}_f \end{bmatrix}, \quad \Xi_{32i} = \begin{bmatrix} X_1^T B + \bar{B}_f \Lambda_i D \\ X_3^T B + \bar{B}_f \Lambda_i D \end{bmatrix},$$

$$\Xi_{43i} = \begin{bmatrix} (\bar{B}_f \Lambda_i \Pi)^T & (\bar{B}_f \Lambda_i \Pi)^T \end{bmatrix},$$

$$\Xi_{33} = \text{diag} \{ \bar{\Xi}_{33}, \bar{\Xi}_{33}, \dots, \bar{\Xi}_{33} \},$$

$$\bar{\Xi}_{33} = \begin{bmatrix} \bar{P}_1 - X_1 - X_1^T & * \\ \bar{P}_2 - X_2 - X_2^T & \bar{P}_3 - X_2 - X_2^T \end{bmatrix},$$

$$\Xi_{51} = \begin{bmatrix} -L & \bar{C}_f \end{bmatrix}, \quad Q \text{ is shown in (9).}$$

Moreover, the filter parameter matrices are given by

$$\begin{bmatrix} A_f & B_f \\ C_f & 0 \end{bmatrix} = \begin{bmatrix} X_2^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \bar{A}_f & \bar{B}_f \\ \bar{C}_f & 0 \end{bmatrix} \quad (14)$$

**Proof:** First we will show the equivalence between the following inequalities and (9)

$$\begin{bmatrix} -P + \lambda \bar{C}^T \bar{C} & * & * & * & * \\ \Phi_{21} & \Phi_{22} & * & * & * \\ \Upsilon_{33}^T \Phi_{33}^{-1} \Phi_{31} & \Upsilon_{33}^T \Phi_{33}^{-1} \Phi_{32} & \Phi_{33} - \Upsilon_{33}^T - \Upsilon_{33} & * & * \\ 0 & 0 & \Phi_{43} \Phi_{33}^{-1} \Upsilon_{33} & -\lambda I & * \\ \tilde{C} & 0 & 0 & 0 & -I \end{bmatrix} < 0 \quad (15)$$

where  $\Upsilon_{33} = \text{diag} \{ W, W, \dots, W \}$

Firstly, if (15) hold, we have (9) from (15) by choosing  $W = W^T = P$  and applying Schur complement. Secondly, it follows by the inequality  $(P - W)^T P^{-1} (P - W) \geq 0$  that  $-W^T P^{-1} W \leq P - W - W^T$ . If we substitute  $P - W - W^T$  by  $-W^T P^{-1} W$  in (15) and multiply them by

$\text{diag}\{I, I, \Phi_{33} \Upsilon_{33}^{-T}, I, I\}$  and  $\text{diag}\{I, I, \Upsilon_{33}^{-1} \Phi_{33}, I, I\}$  on the left and on the right, respectively, then by applying Schur complement we obtain (9). Therefore, (9) and (15) are equivalent.

Let  $W = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix}$  and  $P = \begin{bmatrix} P_1 & * \\ P_2 & P_3 \end{bmatrix}$ . Note that  $X_2$  is invertible, and one can find invertible matrices  $W_3$  and  $W_4$  that  $X_2 = W_3^T W_4^{-1} W_3$

$$\begin{aligned} \text{Let } G &= \begin{bmatrix} I & 0 \\ 0 & W_4^{-1} W_3 \end{bmatrix}, \begin{bmatrix} \bar{P}_1 & * \\ \bar{P}_2 & \bar{P}_3 \end{bmatrix} = G^T \begin{bmatrix} P_1 & * \\ P_2 & P_3 \end{bmatrix} G \\ X_1 &= W_1, \quad X_3 = W_2 W_4^{-1} W_3 \\ \bar{G} &= \text{diag}\{G, I, \hat{G}, I, I\}, \quad \hat{G} = \text{diag}\{G, G, \dots, G\} \\ \begin{bmatrix} \bar{A}_f & \bar{B}_f \\ \bar{C}_f & 0 \end{bmatrix} &= \begin{bmatrix} W_3^T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A_f & B_f \\ C_f & 0 \end{bmatrix} \begin{bmatrix} W_4^{-1} W_3 & 0 \\ 0 & I \end{bmatrix} \end{aligned} \quad (16)$$

Multiplying (15) by  $\bar{G}^T$  on the left and  $\bar{G}$  on the right, respectively and by some matrix manipulations, we obtain (13). Therefore, if (13) hold, then the filtering error system (7) is stochastically stable with an  $H_\infty$  performance  $\gamma$ .

Note that it is still difficult to obtain the filter parameters by the formulation (16). Denote the transfer function of the filter (2) by

$$T_f = C_f(dI - A_f)^{-1} B_f \quad (17)$$

where  $d$  is a unit forward operator. Based on (16) we substitute  $\begin{bmatrix} A_f & B_f \\ C_f & 0 \end{bmatrix}$  in (17) with  $\begin{bmatrix} \bar{A}_f & \bar{B}_f \\ \bar{C}_f & 0 \end{bmatrix}$ . Then it can be obtained (14) by some matrix manipulations and  $X_2 = W_3^T W_4^{-1} W_3$ .

Since (13) is linear in the matrix variables  $X_2, \bar{A}_f, \bar{B}_f$  and  $\bar{C}_f$ , the filter parameters can be obtained by (13) and (14). This completed the proof.

**Remark 4:** Note that (9) and (13) are equivalent so that no conservatism is introduced in the filter design method. Moreover, (13) are linear not only in the matrices variables but also in  $\gamma^2$ . Thus optimal  $H_\infty$  performance can be obtained by solving the following minimization problem

$$\begin{aligned} \min \quad & r = \gamma^2 \\ \text{subject to} \quad & (13) \end{aligned} \quad (18)$$

If it has a solution  $r^*$ , then the optimal  $H_\infty$  performance is  $\gamma^* = \sqrt{r^*}$ .

#### IV. ILLUSTRATIVE EXAMPLE

Consider the plant in [18], where the system matrices are given as follows

$$\begin{aligned} A &= \begin{bmatrix} 0.9617 & 0.0191 & 0.1878 & 0.0012 \\ 0.0370 & 0.9629 & 0.0025 & 0.1789 \\ -0.3732 & 0.1853 & 0.8678 & 0.0179 \\ 0.3528 & -0.3553 & 0.0357 & 0.7840 \end{bmatrix}, \\ B &= \begin{bmatrix} 0.0193 & 0.0373 & 0.1903 & 0.3602 \\ 0 & 0 & 0 & 0 \end{bmatrix}^T, \\ C &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0.1 & 0 \\ 0.1 & 0 \end{bmatrix}, \\ L &= \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (19)$$

It can be seen that  $m = 2$  in this system. In what follows, we first show the filtering performance by applying the proposed method. Choose  $\sigma_1 = \sigma_2 = 0.5$ ,  $\mu = 0.2$  and  $\theta_1 = \theta_2 = \theta = 0.9$ . Then it follows that

$$\varepsilon_{12}^{(1)} = 0.3, \quad \varepsilon_{12}^{(2)} = 0.7, \quad \varepsilon_{21}^{(1)} = 0.7, \quad \varepsilon_{21}^{(2)} = 0.3$$

By solving the minimization problem (18) it can be obtained that the optimal  $H_\infty$  filtering performance is  $\gamma^* = 0.8577$ , and the filter matrices parameters are given by

$$\begin{aligned} A_f &= \begin{bmatrix} -0.0484 & -0.0441 & 0.0467 & 0.0282 \\ -0.6318 & 0.2808 & -0.0059 & 0.1159 \\ -1.0651 & 0.2113 & 0.3189 & 0.1550 \\ -0.9915 & 0.0643 & -0.0202 & 0.5248 \end{bmatrix}, \\ B_f &= \begin{bmatrix} -0.6682 & -0.9814 & -0.6818 & -0.7425 \\ -0.4449 & -0.6559 & -0.4586 & -0.4986 \end{bmatrix}^T, \\ C_f &= \begin{bmatrix} -1.7551 & -0.4284 & -0.4576 & 0.0846 \end{bmatrix} \end{aligned}$$

In the simulation,  $w(k)$  is chosen as a white sequence and a sine function is used to represent the uncertain part of probabilities. The figures for the generated  $w(k)$  and  $\rho(k)$  are omitted to save page. The state trajectories of  $z(k)$  and  $z_f(k)$  are shown in Fig. 4 and the filtering error is shown in Fig. 5. By calculation it can be obtained that

$$\sqrt{\sum_{i=0}^{100} \|e(i)\|^2} / \sqrt{\sum_{i=0}^{100} \|w(i)\|^2} = 0.2270 < \gamma^*$$

showing the validity of the proposed results.

Next we consider the effects of the bound of the uncertainty and the quantisation density on the  $H_\infty$  filtering performance. Choose  $\theta = 0.9$  and solve the minimization (18) by tuning the value of  $\mu$ . The results are shown in Table I, from which we can see larger bound results in poorer  $H_\infty$  filtering performance. Choose  $\mu = 0.2$  and solve the minimization (18) by tuning the value of  $\theta$ . The results are shown in Table II, from which we can see coarser quantiser leads to poorer  $H_\infty$  filtering performance. It can be seen that the simulation results coincide with the results in, for example, [6] and [16].

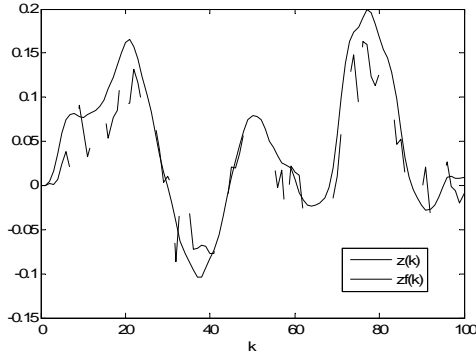


Figure 2. Trajectories of  $z(k)$  and  $z_f(k)$

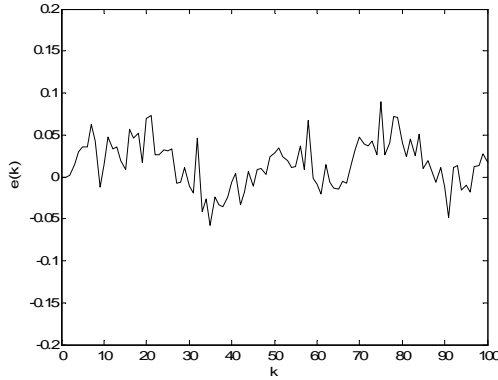


Figure 3. Trajectories of the filtering error  $e(k)$

TABLE I.  $H_\infty$  FILTERING PERFORMANCE WITH DIFFERENT UNCERTAINTY BOUND  $\mu$

$\mu$	$H_\infty$ filtering performance $\gamma^*$
0	0.7993
0.2	0.8577
0.4	0.9277

TABLE II.  $H_\infty$  FILTERING PERFORMANCE WITH DIFFERENT QUANTISATION DENSITY  $\theta$

$\theta$	$H_\infty$ filtering performance $\gamma^*$
0.9	0.8577
0.5	1.9274
0.1	5.5552

## V. CONCLUSION

In this paper, the  $H_\infty$  filtering problem was studied for networked system with uncertain accessing probabilities and quantisation effects. The filtering error system was modeled as a stochastic system with uncertainties both in the system matrix and stochastic parameter. Sufficient conditions are presented for the

stochastic stability with an  $H_\infty$  performance of the filtering error system. A filter design method was also given. An illustrative example was finally given to show the effectiveness of the proposed method. It showed that larger uncertainty bound and coarser quantiser both result in poorer  $H_\infty$  filtering performance.

## REFERENCES

- [1] S. Graham, G. Baliga and P.R. Kumar, "Abstractions, architecture, mechanisms, and a middleware for networked control," IEEE Trans. Autom. Contr., vol. 54, pp. 1490-1503, 2009.
- [2] J.P. Hespanha, P. Naghshtabrizi and Y.G. Xu, "A survey of recent results in networked control systems," Proceedings of the IEEE, vol. 95, pp. 138-162, 2007.
- [3] P.V. Zhivoglyadov, R.H. Middleton, "Networked control design for linear systems," Automatica, vol. 39, pp. 743-750, 2003.
- [4] P. Seiler and R. Sengupta, "An  $H_\infty$  approach to networked control," IEEE Trans. Autom. Contr., vol. 50, pp. 356-364, 2005.
- [5] G.C. Walsh and H. Ye, "Scheduling of networked control systems," IEEE Contr. Syst. Mag., vol. 21, pp. 57-65, 2001.
- [6] M. Fu and L. Xie, "The sector bound approach to quantized feedback control," IEEE Trans. Autom. Contr., vol. 50, pp. 1698-1711, 2005.
- [7] R. Brockett, "Stabilization of motor networks," in 34th IEEE Conference on Decision and Control, New Orleans, USA, 1995, pp. 1484-1488.
- [8] H.J. Gao and T.W. Chen, " $H_\infty$  estimation for uncertain systems with limited communication capacity," IEEE Trans. Autom. Contr., vol. 52, pp. 2070-2084, 2007.
- [9] D. Yue and Q.L. Han, "Network-based robust  $H_\infty$  filtering for uncertain linear systems," IEEE Transaction on Signal Processing, vol. 54, pp. 4293-4301, 2006.
- [10] M. Sahebsara, T.W. Chen and S.L. Shah, "Optimal  $H_\infty$  filtering in networked control systems with multiple packet dropouts," Syst. and Contr. Letters, vol. 57, pp. 696-702, 2008.
- [11] H. Ishii, " $H_\infty$  control with limited communication and message losses," Syst. Contr. Lett., vol. 57, pp. 322-331, 2008.
- [12] M. Tabbara and D. Nesic, "Input-Output stability of networked control systems with stochastic protocols and channels," IEEE Trans. Autom. Contr., vol. 53, pp. 1160-1175, 2008.
- [13] V. Gupta, T.H. Chung, B. Hassibi and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," Automatica, vol. 42, pp. 251-260, 2006.
- [14] P. Seiler and R. Sengupta, "A bounded real lemma for jump systems," IEEE Trans. Autom. Contr., vol. 48, pp. 1651-1654, 2003.
- [15] C.E.de Souza, "Robust stability and stabilization of uncertain discrete-time markovian jump linear systems," IEEE Trans. Autom. Contr., vol. 51, pp. 836-841, 2006.
- [16] J. Xiong, J. Lam, H. Gao and D.W.C. Ho, "On robust stabilization of Markovian jump systems with uncertain switching probabilities," Automatica, vol. 41, pp. 897-903, 2005.
- [17] A.P.C. Goncalves, A. Fioravanti and J.C. Geromel, "Filtering of discrete-time Markov jump linear systems with uncertain transition probabilities," Int. J. Robust. Nonlinear Control, vol. 21, pp. 613-624, 2011.
- [18] H.B. Song, L.Yu and W.A. Zhang, "Networked  $H_\infty$  filtering for linear discrete-time systems," Information Sciences, vol. 181, pp. 686-696, 2011.



# Sliding Mode Control for a miniature helicopter

Jian Fu<sup>1</sup>, Wen-hua Chen<sup>2</sup>, Qing-xian Wu<sup>3</sup>

<sup>1,3</sup>Department of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing, China

<sup>2</sup>Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, UK  
(E-mail: [fujian1986216@126.com](mailto:fujian1986216@126.com), [W.Chen@lboro.ac.uk](mailto:W.Chen@lboro.ac.uk), [wuqingxian@nuaa.edu.cn](mailto:wuqingxian@nuaa.edu.cn))

**Abstract**— A controller based on UAS-SM, sliding mode and open-loop control methodologies is developed for autonomous operation of the Trex-250 helicopter. This controller is composed of a nested sequence of rotor dynamic, angular rate, Euler angle, velocity and position loops. By the controller, coupling of lateral channel and longitudinal channel is greatly reduced. Meanwhile, the feasibility of UAS-SM for implementation is also evaluated in this paper. A practical experiment for Trex-250 is given for the benefits of combined controller.

**Keywords**—UAS-SM; sliding mode; open-loop control; Trex-250; practical experiment

## I. INTRODUCTION

Recently, researchers and industry are interested in the miniature helicopters. Unlike their fixed-wing counterparts, this kind of aircrafts are able to remain stationary over a fixed point, fly backwards and do not require an airstrip to land or take off. Because they are operated in the absence of an onboard pilot, the security is guaranteed. This means, they are more suitable for the tasks like surveillance and reconnaissance in confined and severe surroundings. Although a lot of researchers are engaged in enhancing the autonomy and intelligence of this aircraft, the fundamental problem in the implementation is the autonomous flight control [1].

Because of the highly nonlinear, coupling, underactuated, and inherently instable nature of a miniature helicopter, it is a challenge to design an autonomous flight control system that is capable of operating in the full flight envelope [2].

PID-based feedback control has been proved as an available method for helicopter flight control system [3]. The big advantage of this method is that it can be implemented without a mathematical model and all gains can be tuned by experience in flight test [4]. However, PID control is essentially a linear method. Hence, the achievable performance would be limited when applying classical PID control for miniature helicopter.

Fuzzy logic and neural network have been studied in miniature helicopters in the last years [5]. These algorithms are also non-model-based methodologies. But they possess the capability of approximating nonlinearities and adaptation. Pure fuzzy logic and neural network cannot achieve good performance with regard to unmanned helicopter flight control as the controller would not become fast enough to catch up with practical helicopter. Hence, it is a common way to integrate fuzzy logic and neural network with other methodologies.

Sliding mode variable structure control is considered as a model-based control methodology. Because of its high robustness, miniature helicopters become resistant to the disturbance [6,7]. A big issue of sliding mode algorithm would be the chattering phenomenon. Hence, boundary layer and high-order sliding mode methodologies have been proposed to reduce chattering [8, 9]. But its calculating complexity should be considered, because it might cause the real-time problem.

Furthermore, model-based control methodologies which are usually studied in flight control system of miniature helicopters include multi-model control [10], model predictive control (MPC) [11], linear quadratic (LQ) optimal control [12]. But only a few design methodologies are acceptable when considering practical conditions.

This paper aims to design an efficient controller for Trex-250 helicopter. According to the practical conditions, this controller would be designed by sliding mode with unidirectional auxiliary surfaces (UAS-SM) [14], sliding mode control and open-loop control. It seems that the performance of helicopter is improved with this combined controller by comparing with PID controller. The rest of this paper is designed as follow. Section II contains the system model formulation and experiment setup. Section III details control system design. Section IV gives the analysis and experimental results. And Section V presents some conclusions and future works.

## II. MATHEMATICAL MODEL AND EXPERIMENT SETUP

### A. Mathematical model

The Trex-250 helicopter used in this research is shown in Fig.1. It is a small sized helicopter with the main rotor diameter of 460mm and the trail rotor diameter of 108mm. The belt-driven tail and collective pitch rotor make it capable of 3D maneuvers such as inverted flight. It means that it is well-suited for indoor flight test.



Figure 1. Trex-250 Helicopter

The nested controller is developed with mathematical model obtained from [1]. For the application to Trex-250 we use differential equations as given in (1)-(6). It is obviously that these equations can fall into five different loops, which are position loop (1); velocity loop (2); Euler angle loop (3); angular rate loop (4); and rotor dynamic loop (5). According to these loops, a nested controller is proposed. The detail information of math model is expressed as follow.

$$\begin{aligned} dx/dt &= (\cos\theta\cos\psi) u + (\sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi) v + \\ &\quad (\cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi) w \\ dy/dt &= (\cos\theta\sin\psi) u + (\sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi) v + \\ &\quad (\cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi) w \\ dz/dt &= -(\sin\theta) u + (\sin\phi\cos\theta)v + (\cos\phi\cos\theta)w \end{aligned} \quad (1)$$

$$\begin{aligned} du/dt &= vr - wq - g \sin\theta + X_u u + (T/m) a \\ dv/dt &= wp - ur + g\cos\theta \cdot \sin\phi + Y_v v + (T/m) b \\ dw/dt &= uq - vp + g\cos\theta \cdot \cos\phi + T/m \end{aligned} \quad (2)$$

$$\begin{aligned} d\psi/dt &= p + (\sin\phi\tan\theta) q + (\cos\phi\tan\theta) r \\ d\theta/dt &= (\cos\phi) q - (\sin\phi) r \\ d\phi/dt &= (\sin\phi/\cos\theta) q + (\cos\phi/\cos\theta) r \end{aligned} \quad (3)$$

$$\begin{aligned} dp/dt &= L_a a + L_b b \\ dq/dt &= M_a a + M_b b \\ dr/dt &= N_r r + N_{col}\delta_{col} + N_{ped}\delta_{ped} \end{aligned} \quad (4)$$

$$\begin{aligned} da/dt &= -q/a/T + (A_{lat}/T) \delta_{lat} + (A_{lon}/T) \delta_{lon} \\ db/dt &= -p/b/T + (B_{lat}/T) \delta_{lat} + (B_{lon}/T) \delta_{lon} \end{aligned} \quad (5)$$

$$T/m = Z_w w + Z_{col} \delta_{col} - g \quad (6)$$

where  $x, y, z$  are position along ground axis;  $u, v, w$  are velocities along body axis;  $\phi, \theta, \psi$  are Euler angles;  $p, q, r$  are angular rates;  $a, b$  are flapping angles. The inputs are lateral cyclic ( $\delta_{lat}$ ), longitudinal cyclic ( $\delta_{lon}$ ), tail rotor ( $\delta_{ped}$ ), and main rotor ( $\delta_{col}$ ). Coefficients for this model are shown in TABLE I. This identified model has successfully served the development of a model predictive controller for helicopter autonomous flight [15].

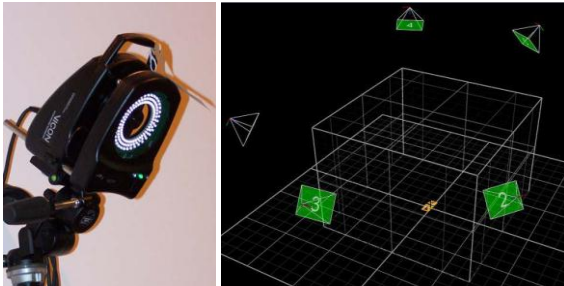


Figure 2. Flight test environment

TABLE I. COEFFICIENTS FOR MATHEMATICAL HELICOPTER

Parameter	Identified value	Parameter	Identified value
$X_u$	-0.233	$Y_v$	-0.329
$Z_w$	-0.878		
$L_a$	83.98	$L_b$	745.67
$M_a$	555.52	$M_b$	11.03
$T$	0.045	$N_r$	-23.98
$A_{lat}$	0.196	$A_{lon}$	1.945
$B_{lat}$	2.120	$B_{lon}$	-0.38
$Z_{col}$	-5.71		
$N_{col}$	8.89	$N_{ped}$	113.65

### B. Experiment setup

Due to the limited payload of Trex-250 helicopter the use of onboard controller hardware was impractical. Instead, the helicopter was controlled by a desktop PC connected to a standard radio transmitter. To provide feedback for the controller, a Vicon Motion Capture system was used. This system includes eight cameras (Fig.2) that can cover a test volume of  $5m \times 4.5m \times 2m$  allowing enough room for Trex-250 maneuvering. With the knowledge of the relative positions of each camera and reflective ball attached to helicopter (Fig.1), Vicon system can determine the position and orientation of the helicopter. Furthermore, the parameter values of  $p, q, r$  are estimated by Simulink block with orientation information.

### III. CONTROL SYSTEM DESIGN

The architecture proposed in this paper is mainly based on the nested three loops in [16]. But some modifications are made because of the practical conditions. The flapping angles  $a, b$  are immeasurable for Vicon system. Meanwhile, it is more convenient for sliding mode design if subsystem is a three dimension system. Therefore, the attitude loop (Fig.4) is divided into three loops which are Euler angle loop (3); angular rate loop (4); and rotor dynamic loop (5). Hence, the architecture is composed of five loops as show in Fig.3 and Fig.4.

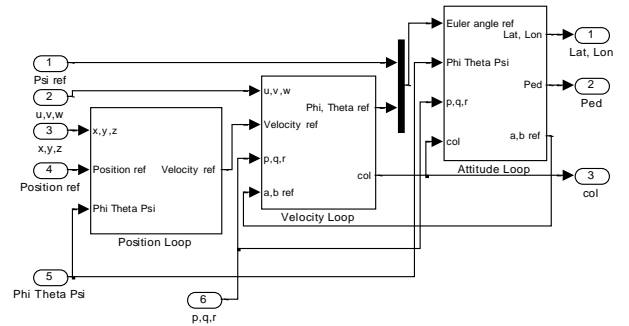


Figure 3. Nested loops

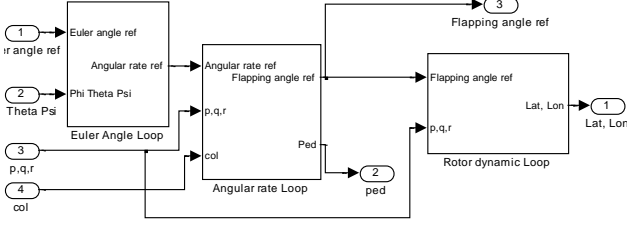


Figure 4. Architecture of Attitude Loop

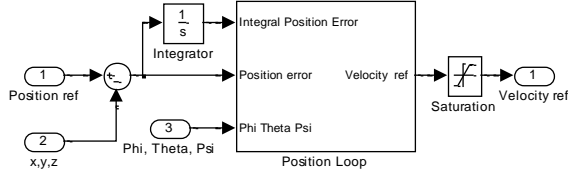


Figure 5. Position Loop

#### A. Position Loop

The position loop is shown in Fig.5. Equations for this loop are given in (1). In this subsystem, position  $x,y,z$  are given as states; velocity  $u,v,w$  are given as control input. It means that designed controller should send the reference velocity  $u,v,w$  to next loop.

The basic UAS-SM equations (see [14]) for a nonlinear system  $dX/dt=f(X)+g(X)u$  are given by the following.

If current unidirectional auxiliary surfaces are defined as

$$H=\Omega_1 X+\Omega_2 \int X+M \quad (7)$$

Then, the control input  $u$  is expressed as

$$u=(g(X))^{-1}(\Omega_1^{-1}N-f(X)-\Omega_1^{-1}\Omega_2 X) \quad (8)$$

where  $\Omega_1=\text{diag}\{\omega_{11},\omega_{21},\omega_{31}\}; \Omega_2=\text{diag}\{\omega_{12},\omega_{22},\omega_{32}\}; H=[H_1;H_2;H_3]'; X=[x;y;z]'; N=[N_1;N_2;N_3]'$  is robust item,  $N \geq 0$ . The symbol  $'$  is given to represent the transpose. And  $\int x$  represents  $\int x dt$  for the sake of simplicity.

Coefficients used in channel  $y,z$  are same with the ones in channel  $x$ . Hence, only  $x$  channel is discussed here. The switching surfaces for  $x$  channel are given as

$$S_{p1}=x+2\int x; S_{p2}=x+\int x \quad (9)$$

TABLE II. COEFFICIENT FOR CURRENT UNIDIRECTIONAL AUXILIARY SURFACE  $H_1$

$\omega_{11}$	$\omega_{12}$	$M_1$	Switching surfaces
0.8	1	1.5	$S_{p1}<0, S_{p2}<0$
-0.4	-0.2	4.5	$S_{p1}<0, S_{p2}\geq 0$
0.4	0.2	4.5	$S_{p1}\geq 0, S_{p2}<0$
-0.8	-1	1.5	$S_{p1}\geq 0, S_{p2}\geq 0$

Current unidirectional auxiliary surface  $H_1$  for  $x$  channel is expressed as follow:

$$H_1=\omega_{11}x+\omega_{12}\int x+M_1 \quad (10)$$

where coefficients are shown in TABLE II, robust item  $N_1=\tanh(0.1\{S_{p1} S_{p2}\})$ . The reference velocities are limited to  $-0.5\text{m/s}\sim 0.5\text{m/s}$  by the saturation block in Fig 5.

Current unidirectional auxiliary surfaces for channel  $y$  and  $z$  can be designed in the same way. UAS-SM is proposed for system with state constraints in [14]. But this constrained effect is not very obvious, because UAS-SM method is not utilized in the rest loops. Experiments for the constrained effect of UAS-SM would be done in the further research. Hence, only the stability of UAS-SM is tested here.

#### B. Velocity Loop

The velocity loop is shown in Fig.6. Equations for this loop are given in (2). The subsystem is utilized to track reference velocity  $u,v,w$  from position loop. As shown in (2), velocity  $u,v,w$  are given as states; Euler angles  $\phi,\theta$  and main rotor control  $\delta_{col}$  are given as control inputs. It means that the designed controller should send reference Euler angles  $\phi,\theta$  to the attitude loop.

As discussed before, flapping angles  $a,b$  are immeasurable by Vicon system. The reference flapping angles from Attitude loop are used to estimate  $a,b$ . Because normal Euler angles are no more than 14 deg, small-angle approximation is used to transfer the non-affine system (2) into the affine one. The simplified velocity system is expressed as follow.

$$\begin{aligned} du/dt &= vr-wq-g\cdot\theta+X_u u+T/m a \\ dv/dt &= wp-ur+g\cdot\phi+Y_v v+T/m b \\ dw/dt &= uq-vp+g+T/m \\ T/m &= Z_w w+Z_{col}\delta_{col}-g \end{aligned} \quad (11)$$

The sliding mode surfaces used in this loop is expressed as follow.

$$S_v=K_{v1}\{u,v,w\}'+K_{v2}\cdot[\int u,\int v,\int w]' \quad (12)$$

where  $S_v=[S_v(1),S_v(2),S_v(3)]'$ ,  $K_{v1}=\text{diag}\{0.5,0.5,0.2\}$ ,  $K_{v2}=\text{diag}\{1,1,1\}$ , robust items are given as

$$[-0.1 \tanh(S_v(1)), -0.15 \tanh(S_v(2)), -0.5 \tanh(0.5 S_v(3))]'.$$

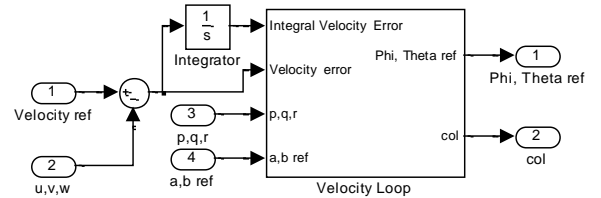


Figure 6. Velocity Loop

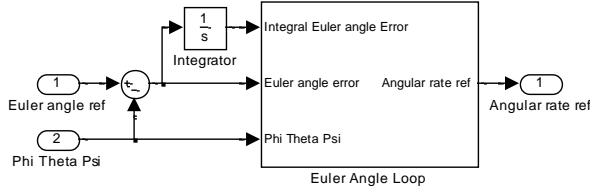


Figure 7. Euler Angle Loop

### C. Euler Angle Loop

The Euler angle loop is shown in Fig.7. Equations for this loop are given in (3). In this subsystem, Euler angle  $\phi, \theta, \psi$  are given as states, reference angular rates  $p, q, e$  are given as control inputs.

A normal sliding mode design is used here. Switching surfaces are expressed as follow.

$$S_E = K_{E1} \{\phi, \theta, \psi\}' + K_{E2} \{[\dot{\phi}, \dot{\theta}, \dot{\psi}]\}' \quad (13)$$

where  $S_E = [S_E(1), S_E(2), S_E(3)]$ ,  $K_{E1} = \text{diag}\{0.1, 0.1, 0.5\}$ ,  $K_{E2} = \text{diag}\{1, 1, 1\}$ , robust items are given as

$$[-0.1 \tanh(S_E(1)), -0.1 \tanh(S_E(2)), -0.1 \tanh(S_E(3))].$$

### D. Angular Rate Loop

Angular Rate Loop is shown in Fig.8. Equations for this loop are given in (4). In this subsystem, angular rates  $p, q, r$  are given as states, tail rotor ( $\delta_{ped}$ ) and flapping angle references are given as control inputs.

A sliding mode controller which is similar to the one in Euler angle loop is expressed as follow.

$$S_A = K_{A1} \{p, q, e\}' + K_{A2} \{[\dot{p}, \dot{q}, \dot{r}]\}' \quad (13)$$

where  $K_{A1} = \text{diag}\{0.04, 0.04, 0.04\}$ ,  $K_{A2} = \text{diag}\{1, 1, 1\}$ ,

$S_A = [S_A(1), S_A(2), S_A(3)]'$ , robust items are given as

$$[-\tanh(0.01 * S_A(1)), -\tanh(0.01 * S_A(2)), -0.1 * \tanh(S_A(1))].$$

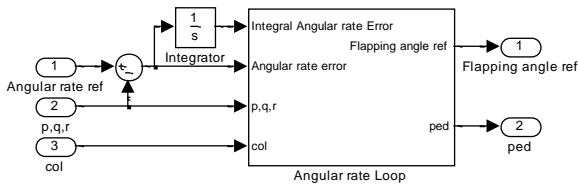


Figure 8. Angular rate Loop

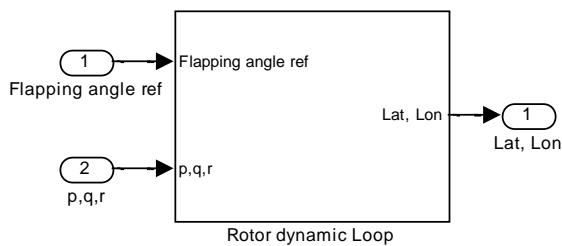


Figure 9. Rotor Dynamic Loop

### E. Rotor Dynamic Loop

The controller for rotor dynamic loop is a little different from other loops in this paper, because the states of rotor dynamic system (5) are immeasurable from Vicon system. It is possible to obtain the flapping angles by state observer. But this method will not be discussed in this paper. Hence, a open-loop controller is applied in this loop as shown in Fig.9.

Normal range of flapping angle  $a, b$  would be  $-0.05 \sim 0.05$ . It can be considered as a small value for equations. Hence, the derivations of flapping angle are assumed to be zero.

$$da/dt \approx 0; \quad db/dt \approx 0 \quad (14)$$

Hence, the control inputs lateral cyclic ( $\delta_{lat}$ ), longitudinal cyclic ( $\delta_{lon}$ ) are expressed as

$$[\delta_{lat}, \delta_{lon}]' = ([A_{lat}/T \quad A_{lon}/T; B_{lat}/T \quad B_{lon}/T])^{-1} [q + a/T, p + b/T]' \quad (15)$$

## IV. ANALYSIS AND EXPERIMENT RESULTS

The performance of Trex-250 with combined controller and PID controller is shown in Fig. 10 ~Fig.12. Europe coordinate system is used in this paper. Therefore, it is positive downwards for height direction. There are two vertical lines annotated in Fig. 10~Fig.12. The left vertical line indicates the time at which helicopter is taking off. And the right vertical line correspond to the time when helicopter is landing down.

From Fig.11, the dynamic system of Trex-250 with UAS-SM method in position loop is stable. It implies that the UAS-SM method is available for the implementation of helicopter. Hence, some further researches with UAS-SM controller would become feasible.

Ground effects are found when helicopter is taking off as shown in Fig. 10 and Fig.11. Combined controller gives larger deviations than PID controller. It is because that the mathematical model of helicopter is inaccurate because of the ground effect. For the sake of simplicity, this effect would not be discussed in this paper. Hence, these deviations are insignificant. Actually, they would be reduced by considering the math model with ground effect. It will be investigated in the further research.

A delay for about 10 seconds in Z direction is observed in Fig. 10, when helicopter is taking off with PID controller. It is caused by an offset given to guarantee throttle input for PID would begin with -1. Therefore, this delay should be acceptable.

The reference signal for height is 0.2m as given in Fig. 10 and Fig.11. Based on the past experiments, helicopter would hover at about -0.1m when the reference signal is 0m. Because of the ground effect, the lift of helicopter is higher than what we expected. Hence, helicopter would be forced to land by 0.2m reference signal.

As discussed before, PID control methodology is essentially a linear method. The Z direction and X, Y directions are controlled respectively. But they are coupled in this experiment. Deviation from the X, Y directions might cause a changed lift of main rotor.

Therefore, Z direction of helicopter is suffering about 0.1m error, as shown in Fig.10, when helicopter is tracking a new position with PID controller. It is shown as a visible up-and-down action during the experiment.

From Fig.11, it is obvious that combined controller gives an accurate tracking in the Z direction. The error between reference signal and practical height of helicopter is about 0.04m. The UAS-SM and sliding mode used in combined controller are nonlinear control methodologies. Based on these methods, the coupled effect between Z direction and X, Y directions are considered.

For security, normal velocity used indoor experiments is about  $\pm 0.5$  m/s, and no more than  $\pm 1$  m/s for the maximum value. It is because that Vicon system might not catch up with helicopter when it is moving with high speed. However, we got some peak values which are -1.5m/s in Fig. 12 when helicopter is moving from 1m to -1m with PID controller. It should be avoided because we may crash the helicopter when Vicon system lost the target. Unlike PID controller, combined controller gives a limited velocity as shown in Fig.12. It seems that the saturation block in Fig.5 is efficient.

## V. CONCLUSION AND FUTURE WORK

A nested controller of Trex-250 with UAS-SM, sliding mode, and open-loop methodologies provides good trajectory tracking reference position signal. This paper is aimed at presenting the different performance between PID and sliding mode. The stability of UAS-SM in practical experiment is also evaluated in this paper. A UAS-SM controller will be implemented and flight tested on Trex-250 to evaluate the constrained effects of this method. Another further research would focus on the reduction of ground effect.

## ACKNOWLEDGMENT

This work was carried out using the Vicon-Nexus tracking systems at Loughborough University.

Thanks to Dr. Wen-hua Chen, Jonathan H.A. Clarke, and Cunjia Liu for their technical support during the experiment.

## REFERENCES

- [1] Cunjia Liu, J. Clarke, Wen-hua Chen, "Modeling and identification of a miniature helicopter", Proceedings of Workshop on Human Adaptive Mechatronics(HAM), 2010.
- [2] Yunjun Xu, "Multi-Timescale Nonlinear Robust Control for a Miniature Helicopter", IEEE Transactions on Aerospace and Electronic systems, Vol. 46, No. 2, pp 656-671, April 2010,.
- [3] S. Bouabdallah, A. Noth and R. Siegwart, "PID vs LQ Control Techniques Applied to an Indoor Micro Quadrotor", Proceedings of 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, September 28-October 2, 2004, Sendai, Japan.
- [4] B. Mettler, "Identification Modeling and Characteristics of Miniature Rotorcraft". Norwell, MA 02061: Kluwer Academic Publishers, 2003.
- [5] Xiaodong Wang, Xiaoguang Zhao, "A Practical survey on the Flight Control System of Small-Scale Unmanned Helicopter", Proceedings of the 7<sup>th</sup> World Congress on Intelligent Control and Automation, Chongqing, China, June 25-27, 2008,.
- [6] Wei Wang and Gang Song, "Autonomous Control for Micro-Flying Robot and Small Wireless Helicopter X.R.B", Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, October, 2006.
- [7] Wei Wang, Kenzo Nonami and Yuta Ohira, "Model Reference Sliding Mode Control of Small Helicopter X.R.B based on Vision", International Journal of Advanced Robotic Systems, pp.235-242, Vol. 5, No.3, 2008.
- [8] Min-Shin Chen, Yean-ren Hwang, and M. Tomizuka, "A state-Dependent Boundary Layer Design for Sliding Mode Control", IEEE Transactions on Automatic Control, Vol 47, No. 10, pp:1677-1681, October 2002.
- [9] H. Rios, A. Rosales, "Robust Regulation for a 3-DOF Helicopter via Sliding-Modes Control and Observation Techniques", 2010 American Control Conference Marriott Waterfront, Baltimore, MD, USA June 30-July 02, 2010.
- [10] D. Godbole, T. Samad and V. Gopal, "Active Multi-Model Control for Dynamic Maneuver Optimization of Unmanned Air Vehicles," in proceedings of the 2000 IEEE International Conference on Robotics & Automation, pp. 1257-1262, April 2000.
- [11] T. Templeton, D. Hyunchul Shim, C. Geyer, and S. Shankar Sastry, "Autonomous Vision-based Landing and Terrain Mapping Using an MPC-controlled Unmanned Rotorcraft", 2007 IEEE International Conference on Robotics and Automation Roma, Italy, 10-14 April, 2007.
- [12] Z. Jiang, J. Han, Y. Wang and Q. Song, "Enhanced LQR Control for Unmanned Helicopter in Hover," in 1<sup>st</sup> International Symposium on Systems and Control in Aerospace and Astronautics, pp. 19-21, January 2006.
- [13] M. G. perhinschi, "A modified Genetic Algorithm for the Design of Autonomous Helicopter Control System," AIAA Guidance, Navigation, and Control Conference, August 1997.
- [14] FU Jian, WU Qing-xian, JIANG Chang-sheng, CHENG Lu, "Robust sliding mode control with unidirectional auxiliary surfaces for nonlinear system with state constraints", Control and Decision, Vol 26, No.7, July 2011, in press.
- [15] C. Liu, W.-H. Chen, and J. Andrews, "Model predictive control for autonomous helicopters with computational delay in consideration", in UKACC 2010.
- [16] Dale Enns, Tamas Keviczky, "Dynamic Inversion Based Flight Control for Autonomous RMAX Helicopter", Proceedings of the 2006 American Control Conference Minneapolis, Minnesota, USA, June 14-16, 2006.

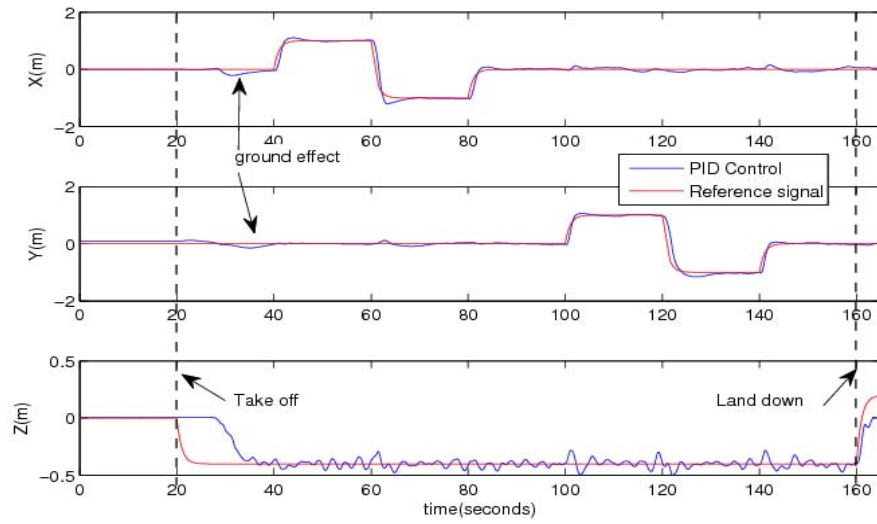


Figure 10. Position response of PID Control

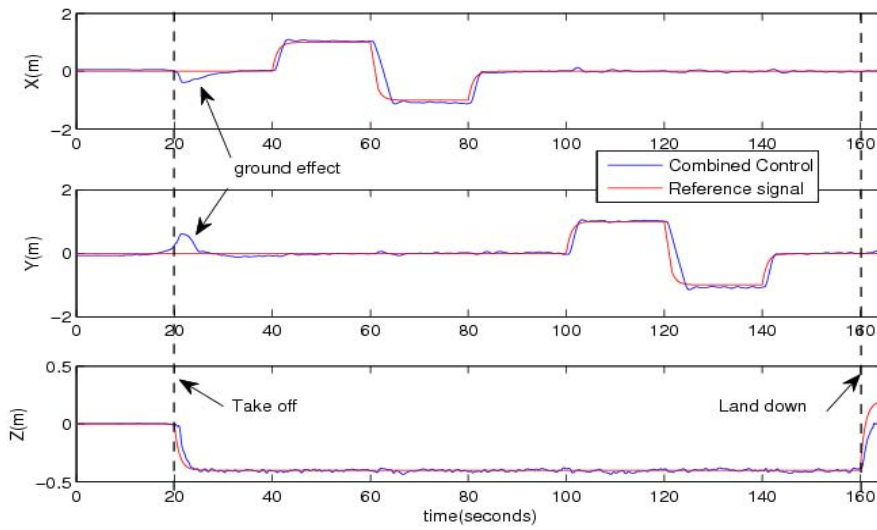


Figure 11. Position response of Combined Control

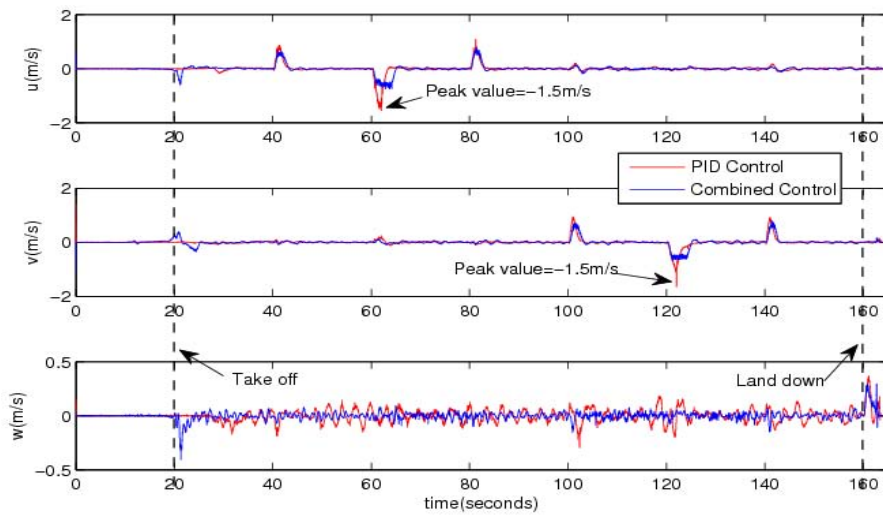


Figure 12. Velocity response of PID and Combined Control



# Real-Time Implementation of a Burst Error Compensator for Wireless Control Systems

Michael Short<sup>1</sup>, Usama Abrar, Ian French and Fathi Abugchem

Electronics & Control Group  
Teesside University  
Middlesbrough, U.K.

<sup>1</sup>Corresponding Author: [m.short@tees.ac.uk](mailto:m.short@tees.ac.uk)

**Abstract**—The use of wireless communications systems in automation and real-time control applications is increasing at a steady rate. The use of wireless technologies in these applications poses several severe problems, which include out-of-order packet transmissions, high levels of packet jitter and high probabilities of packet losses; these problems are especially problematic in systems with strict timing constraints. This paper is concerned with ameliorating the effects of packet loss in time-triggered sampled-data control systems, in which the feedback connection is implemented via one or more wireless links. In particular, the paper develops a simple predictive compensator that reconstructs the best estimates of a sequence of one or more missing data samples using an ARMA model of the process under control. An embedded implementation of the compensator is applied to a case study, in which ARM7 microcontrollers and ISA band wireless transceivers are employed in a real-time servomotor control system experiencing artificially induced packet losses. Experimental results are presented which indicate that the technique has the potential to maintain system tracking performance and stability in the presence of sever interference and burst errors affecting the feedback channel.

**Keywords**—Real-time control, Distributed systems, Wireless sensor/actuator networks, Predictive control.

## I. INTRODUCTION

The use of wireless communications systems in automation and real-time control applications is increasing at a steady rate. Wireless systems have the distinct advantage of reducing equipment installation complexity through the lack of a need for wiring and harnessing, enabling easier trouble-shooting and system re-configuration; this reduces the long-term maintenance requirements associated with wired systems [1-4]. In addition, wireless sensor/actuator networks (WSANs) can potentially provide the device interconnectivity needed for a range of industrial control and monitoring functions across a wide range of operating environments [1-4].

However, wireless systems are generally perceived in a negative sense for real-time applications such as the sensor/actuator networks needed for industrial process control systems [5]. The use of wireless technologies in these applications poses several severe problems, which include out-of-order packet transmissions, high levels of packet jitter and high probabilities of packet losses; these problems are especially problematic in systems with strict timing constraints [4-7]. Although much progress has

been made in recent years, to date most industrial applications of wireless technology have mainly been restricted to soft real-time process monitoring and data acquisition applications, in which interruptions to the wireless service do not lead to unacceptable loss of control or damage to equipment [4][7]. When control loops have been closed by wireless equipment in industrial situations, developers have often been forced to take drastic actions; for example the enforcement of ‘blackout zones’ around the feedback control loops in question [7]. Indeed, as discussed by [4] and [7], the use of wireless technology in feedback control applications presents arguably the largest ongoing challenge in the domain, and is the focus of the current paper.

The particular structure of WSAAN architecture under consideration in this paper is as illustrated in Figure 1, in which an intelligent sensor communicates to an intelligent actuator via one (or more) wireless channels. In the paper, we are concerned with ameliorating the effects of feedback packet loss in such a sampled-data control system. It is assumed that a time-triggered architecture - loosely based upon a TDMA media access control scheme - is employed in the wireless channels, and that the setpoint signal is fully available in the intelligent actuator i.e. it is not affected by packet losses. In particular, the paper will begin to develop a simple predictive compensator based around a recursive Auto Regressive Moving Average (ARMA) process model, which reconstructs estimates of a sequence of one or more missing data samples. Under the assumption that a suitably detailed model of the process is available, these estimates will be optimal in a least squares sense.

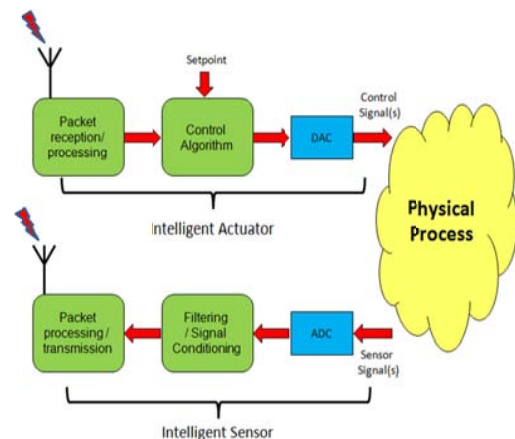


Figure 1: Wireless feedback control system.

The remainder of the article is organized as follows. Section II reviews some previous work in this area, and Section III presents main operation of the proposed predictive compensator. A prototype embedded implementation of the compensator in a small WSN is described in Section IV, which also describes a simple case study and experimental results. After a short discussion, the paper is concluded in Section V.

## II. PREVIOUS WORK

### A. Wireless Communications and Real-Time Control

As stated in the introduction, in this paper we assume that the sampled-data control system to be implemented possesses a time-triggered communications scheme as a part of its underlying system architecture. Time-triggered (TT) communications are thought to increase the predictability and overall reliability of many types of system; evidence would suggest this is especially apparent for real-time control systems [8][9][10]. Basic TT schemes have also found to be of use in WSNs, however hybrid TT schemes - that can help to reduce node power consumption - have found to be better suited in practice [3][11][12]. Conceptually, these techniques tend to rely on variations of the use of Time Division Multiple Access (TDMA) message schedules such as that depicted in Figure 2, in which  $p$  represents the inter-slot spacing.

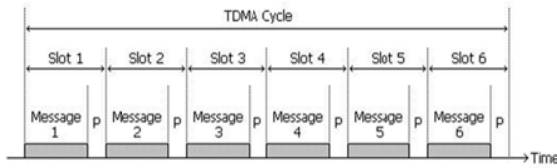


Figure 2: A typical TDMA communications schedule.

Most existing wireless protocols for industrial applications (such as WirelessHART and ISO 100.11a) employ TDMA-based schemes [4]. Although such TDMA-based systems have been found effective to help reduce levels of packet jitter, they provide little resilience against the effects of packet loss and channel dropout. In wired networks such as CAN, Bit Error Rates (BERs) may be in the region of  $10^{-6}$ , arriving in short correlated bursts of between 5-20 bits [13]. As such, packet loss probability in these situations is comparatively low, and the omission of multiple consecutive samples can effectively be considered to be a 'rare' event.

However the nature of a wireless channel is such that packet loss probabilities are comparatively higher, and – importantly – will typically be strongly correlated [14][15]. For example, experimental data have been reported in the literature that shows that wireless channels having Packet Error Rates (PERs) as low as 0.01% will regularly experience losses of over 50 consecutive packets, rising in some cases to over 1,000 [3]. As such the omission of multiple consecutive samples can be considered to be a 'routine' event. In many types of system, the loss of the wireless channel for a duration of time exceeding the feedback controllability limit of the process (closely related to the 'ultimate deadline' as discussed in [16]) may become an occurrence that cannot

be neglected from the design. Whilst off-line routing algorithms [3] and MAC-level enhancements [7] can help to ameliorate these issues, they cannot be completely eliminated and therefore application-level compensation schemes are also sought.

### B. Packet Loss Compensation

Several previous works have considered the compensation of omitted samples in networked systems, both wired and wireless. In the latter case, given the high prevalence of multimedia applications in this domain work has principally concentrated on application-specific receiver interpolation and error correction of lost samples in applications such as streaming audio and video – see [17] for an overview. In the former case, a larger amount of work related to control systems may be found. This work ranges from relatively simple first-order hold interpolators to be used in intelligent actuators ([18]) through to complex techniques based upon  $H_\infty$  filtering of the data stream ([19][20]). Whereas the simple techniques are trivial to implement, they are principally aimed at compensation for the loss of single control signal samples at irregular intervals and would seem to be less suitable to the wireless domain for the reasons discussed above. Although the more complicated techniques would seem to be much better suited, their implementation is non-trivial; in both cases, [19] and [20] require both a high-fidelity process model, and the off-line solution to large non-convex (and hence non-linear) optimization problems. In addition, both [19] and [20] assume that samples losses are independent and uncorrelated. In the following Section, we will introduce a midway technique that is conceptually much easier to implement, and should provide reasonable resilience to the effect of burst errors in wireless feedback channels.

## III. PREDICTIVE COMPENSATION OF PACKET LOSS

In this Section, the predictive compensator is introduced. It is assumed in this initial study that the system has a topology as illustrated in Figure 1, and that the sensor node attempts to transmit data packets with a regular period, at appropriate slots in the TDMA cycle. We also make the simplifying assumptions that the clocks between sensor and controller/actuator nodes are synchronized using an appropriate clock synchronization algorithm, and that the level of clock drift that may occur relative to any continuous interval of network unavailability is small enough to be neglected.

### A. General Procedure

The predictive compensator proposed in this paper consists of three main elements, the arrangement of which inside the intelligent actuator/controller is as shown in Figure 3. The schedule decoder/packet loss detector provides the indication of missing samples by polling to check for the presence (or absence) of new data packets following the expiry of the appropriate slot(s) in the TDMA cycle. In the presence of a fresh sample, the packet is processed accordingly and the signal directly forwarded to the controller (via the switch), prior to the execution of the control algorithm. When an omission is detected, a signal is asserted to the switch, such that an estimate of the omitted packet's content is instead

forwarded to the controller. This estimate is produced by the ARMA interpolator. The interpolator continually uses a technique - to be described below - to predict the values of samples, by monitoring a (usually short) time history of the control signal  $u(k)$  and the actual/estimated process values  $y(k)$  and  $\hat{y}(k)$  respectively. The nature of the ARMA model is further described in the next Section.

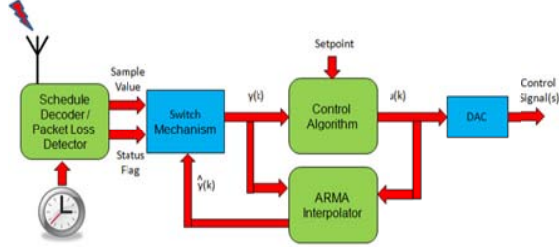


Figure 3: Structure of the packet loss compensator.

### B. ARMA Process Model

In the design of a sampled-data control system, a nominal model of the process to be controlled will normally be identified by the control engineers at an early stage of the design process. This model may be either continuous or discrete in nature; conversion from the former to the latter can be achieved by using the  $z$ -transform. As such, it is assumed that the process under control can be described by the discrete-time linear pulse transfer function:

$$\frac{Y(z^{-1})}{U(z^{-1})} = z^{-d} \frac{B(z^{-1})}{A(z^{-1})} \quad (1)$$

Where  $z^{-1}$  is the backward shift (delay) operator,  $U$  and  $Y$  represent the discrete process input and output, and with  $A$  and  $B$  having the following polynomial definitions:

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n} \\ B(z^{-1}) &= b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_m z^{-m} \end{aligned} \quad (2)$$

Note that in (1),  $d$  represents the integer part of the system time-delay such that  $b_0$  is non-zero. Assuming that a zero order hold (or similar) is employed for control purposes,  $d$  itself will also be non-zero in a causal system. The transfer function model described by equations (1) and (2) is representative of many types of real-world physical processes, see for example [21][22][23]. In terms of creating a process model which may be used for short-term predictions of the process output, equation (1) may be solved for  $Y$  and via the definition of the delay operator may be transformed into an Auto-Regressive Moving Average (ARMA) difference equation given by (3). In this equation, the  $a$  and  $b$  coefficients have identical values to those of (2) – with the signs of the  $a$  coefficients reversed – and with  $u(k)$  and  $y(k)$  representing the discrete process input and output at sample index  $k$ , where  $t = kt_s$  for a sampling time of  $t_s$  seconds.

$$\begin{aligned} y(k) &= a_1 y(k-1) + \dots + a_n y(k-n) \\ &+ b_0 u(k-d) + \dots + b_m u(k-d-m) \end{aligned} \quad (3)$$

### C. Predictive Packet Loss Compensation

Assuming that the time history of the process input and output is known at sample instant  $k$ , then equation (3) may be used to predict the process output at sample  $k+1$ ; indeed this forms the basis of techniques such as model predictive control [22]. However, in this paper the model is employed to predict the *current* process output  $y(k)$  in situations in which a packet is omitted; the known values of  $u(k) \dots u(k-d-m)$  are employed along with the known (and possibly *previously* predicted) values of  $y(k) \dots y(k-n)$ . As such, in situations in which many consecutive samples are omitted, the controller applies a form of open-loop control to drive the process towards the setpoint. Assuming that the model is of reasonable quality – and the duration of the interference is not too excessive – the compensator should achieve a reasonable quality of control and retain a straightforward structure with a simple code implementation. In the next Section, we consider a case study using an embedded implementation of the predictor.

## IV. CASE STUDY

In this Section, we will first describe an embedded implementation of a prototype real-time WSN network, consisting of multiple embedded processors and RF transceivers, followed by the case study.

### A. Hardware Configuration

The main processing element in the node hardware platform is based around an LPC2387 microcontroller from NXP semiconductors. This microcontroller has a 32-bit ARM7TDMI-S core, and is a typical hardware platform for the implementation of modern embedded real-time systems. The device can be operated with a CPU clock speed of up to 72 MHz, has 512 Kb of on-chip flash, 98 Kb of on-chip RAM, and a rich set of I/O peripherals. The latter includes multiple CAN Controllers and UARTs, USB and Ethernet support, Multiple Timers, PWM, SPI and Analog/Digital I/O. In order to implement the wireless communication links, a set of RF serial communication transceivers from USBscope were employed. These devices work at 433 MHz, have a range of 300 m and implement half-duplex RS-232/485 like links at each end, operating at 57,600 bps. As such, they may interface directly to an on-chip UART of the LPC2378 and implement point-to-point or multicast communications, with hardware support for error detection. Automatic retransmission in case of error is not employed, and can be implemented with a higher level protocol. Packet sizes of up to 80 bytes may be employed. For simplicity, in this study packets of 16 bytes were used as this is the size of the on-chip UART FIFO buffers, and hence an entire packet can be stored in each TX and RX buffer and transmission and reception can be handled concurrently to the CPU activity, creating a simple but flexible solution with low CPU overheads. The hardware components employed are shown in Figure 4. Two nodes having a configuration as shown in Figure 1 were employed in the case study.

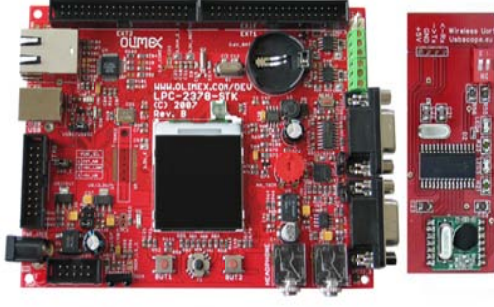


Figure 4: Embedded hardware showing LPC2378 Processor (left) and wireless transceiver (right).

### B. Task and Message Scheduling

A simple means to synchronize the clocks and control the scheduling and run-time execution of application level tasks was employed for this prototype system. Each of the embedded nodes runs a task scheduler, which is driven by a local timer. Preemptive Earliest Deadline First (EDF) scheduling is used to control CPU task executions, which are all periodic in nature. A wireless variant of a shared-clock scheduler [8] was employed to control message transmissions and receptions, and implement TDMA scheduling. Each 'Tick' message from the designated master node (actuator node in Figure 1) is transmitted with a 16-bit timestamp, enabling the slave node (sensor node in Figure 1) to adjust its local clock and synchronize task executions in the distributed system. Although this simple approach is not yet robust enough for industrial applications, it provides a platform that is reliable enough for laboratory experimentation. In the next Section, the process that was controlled in the case study is described.

### C. Process Description

In real-time control systems, there are several procedures to assist with the selection of a suitable sampling rate, see e.g. [22][23]. Assuming a suitable sampling rate has been selected, the behavior of the closed loop system should not be affected too much as a result of a single dropped sample, regardless of the bandwidth or speed of the process dynamics. However, for systems with comparatively fast dynamics, a faster sampling rate will generally be needed; in turn, the length of time between samples - and hence chance for packet re-transmission in case of error - is shortened. As the ultimate deadline will also be comparatively shorter, intuitively it is expected that systems with faster dynamics will be more susceptible to burst errors. For this reason, in this case study a servomotor with simple (but comparatively fast) dynamics was chosen as the process to control. The servomotor (supplied by Feedback Instruments©) is as shown in Figure 5. For simplicity reasons, it was decided to control the position of the motor using a digital PV (Proportional Velocity) controller. Velocity feedback was employed to apply damping as opposed to derivative action in order to reduce the effects of setpoint 'kick'. The PV controller was tuned to give an approximate phase margin of 60°, and a step response settling time of  $\approx 2$ s. A sampling time of 100ms was employed for the control design.

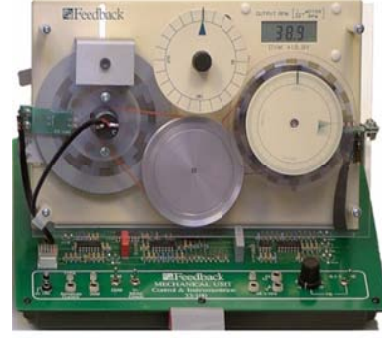


Figure 5: Servomotor trainer employed in the case study.

Considering the relationship between applied DC voltage and shaft position, the type-I system dynamics can be represented by a first order lag plus integrator. The pulse transfer function of the servomotor (after taking in account the effects of a ZOH, plus an additional delay of one sample due to communication latency in the feedback path) was determined experimentally to be as follows:

$$\frac{Y(z^{-1})}{U(z^{-1})} = z^{-2} \frac{0.01873 + 0.01752z^{-1}}{1 - 1.818z^{-1} + 0.818z^{-2}} \quad (4)$$

The identification was performed using offline parameter estimation, in which the open loop system was excited with a square wave input for a short time duration. The ARMA Instrumental Variables algorithm in the Matlab© system identification toolbox was employed for this identification. A linear parametric model with structure  $n = m = 2$  and a fixed delay  $d = 2$  was employed, and the approximated mathematical model gave an accuracy of fit of 94%. The accuracy of fit statistic is a normalized coefficient-of-determination, representing the proportion of the input/output dynamic response that is explained by the fitted regressors in the model, see [21][22]. From this model, the code required to implement the predictive compensator could easily be determined. When a packet is omitted, equation (5) is used to reconstruct the missing sample at index  $k$ :

$$\hat{y}(k) = 0.01873u(k-2) + 0.01752y(k-3) + 1.818y(k-1) - 0.818y(k-2) \quad (5)$$

Where  $y(k)$  becomes  $\hat{y}(k)$  in the case of missing samples, and this estimate becomes the value that is passed to the control algorithm and subsequently used in the recursion if sample  $y(k+1)$  is also missing. The predictor was coded into an application task on the master node, and equation (5) invoked whenever a missing sample was detected.

### D. Experimental Setup

In a controlled laboratory setting, as may be expected initial experimental data indicated that errors were few and far between. In order to simulate the effects of burst



errors, some simple fault injection techniques were therefore employed. The most common way to model bursty behavior in communication links is to use a simple two-state Markov model of the Gilbert or Gilbert-Elliot types [14][15], the latter of which is shown in Figure 6.

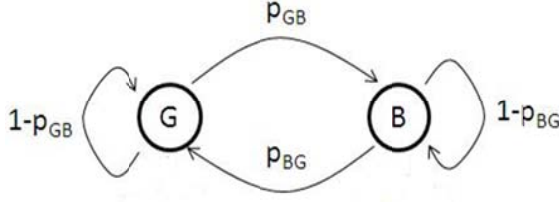


Figure 6: Two state Gilbert-Elliot model of a bursty communication link.

The model has two states  $G$  and  $B$ , representing ‘Good’ and ‘Burst’ states respectively. Transitions between the two states  $G$  and  $B$  have associated with them the probabilities  $p_{BG}$  and  $p_{GB}$ . The probability of remaining in a given state is then given by  $p_{GG} = 1 - p_{GB}$  and  $p_{BB} = 1 - p_{BG}$ . Each state also has associated with it a probability of bit error occurrence ( $\beta_G$  and  $\beta_B$ ). Although this may be useful for exploring the effectiveness of error detection schemes such as CRCs, for simplicity these probabilities can be set to 0 and 1. The model parameters  $p_{GB}$  and  $p_{BG}$  can be loosely interpreted as follows: the reciprocal of  $p_{GB}$  defines a mean error gap length  $\mu_{EG}$ , and the reciprocal of  $p_{BG}$  defines the mean error burst length  $\mu_{EB}$ , both having a geometric distribution. In this study, a Gilbert-Elliot model was used to deliberately drop ‘Ack’ packet transmissions containing sensor data in the slave node. A mean error gap length of 50 packets and a mean error burst length of 10 packets were employed, with random numbers being used to update the model state at each sample interval. Note that the mean error gap was deliberately shortened, such that multiple bursts would be encountered in short experimental runs in the laboratory.

#### E. Experimental Results

Two experiments were carried out, one with and one without the compensator. In each experiment, the control system was required to track an input square wave of amplitude 10v and period 20s in the presence of burst errors. Packet loss was suppressed for the first 10s to allow the system to stabilize, and each experiment was conducted for a total of 100s. Figure 7 shows the results of the system performance without the predictive compensator. The upper section of the figure shows the process output, whilst the middle section shows the control signal. Also indicated on the lower section is an indicator of burst errors, the line having a value of 1 when an error occurs and 0 otherwise. The effect of packet loss is self-evident; with the process in steady-state and the burst duration relatively short, the data loss has little effect on tracking performance. However when the process is in transient and packet losses occur, the system performance becomes seriously degraded. This is illustrated after  $t = 80$  seconds when the setpoint is overshoot by over 100%, and control is only restored when the burst subsides. The loss of control is also reflected in the control output, with the control signal effectively freezing shortly after the burst occurs.

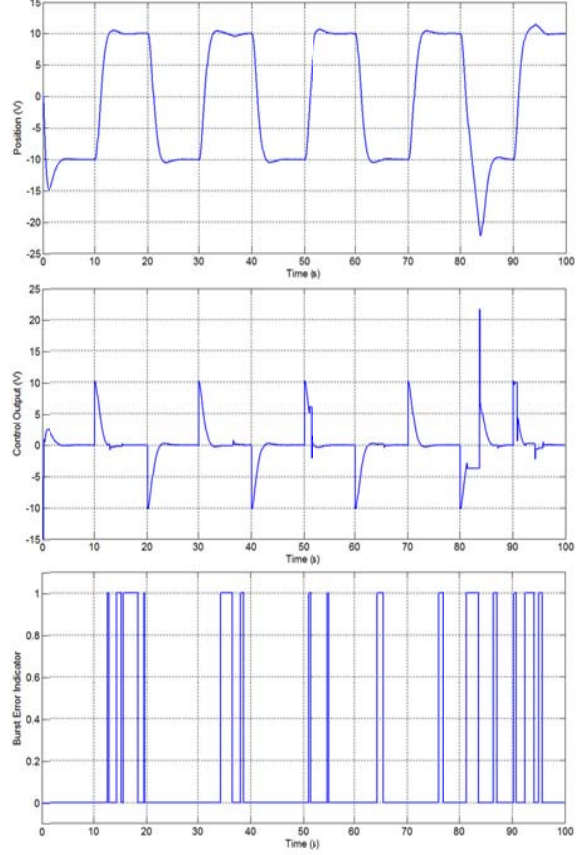


Figure 7: Effect of packet loss without compensation - refer to text for description.

Figure 8 shows the results of the system performance with the predictive compensator in the loop, and the Figure has a similar layout to that of the previous. This time, despite the occurrence of multiple bursts – again during transient process behavior – the tracking performance of the system is maintained. This is best illustrated by considering the step change that occurs at  $t = 60$ s, during a burst error. Despite the complete loss of process information, the compensator is able to interpolate the process predicted response, such that the desired closed-loop performance is maintained. The disturbance rejection properties of the controller may still be affected during a burst, but this was not considered in the study.

#### V. CONCLUSIONS AND FURTHER WORK

This paper has considered a predictive compensation scheme to help ameliorate the effects of packet loss and burst errors in WSANs, and has considered its embedded implementation in a simple real-time control system. Preliminary results suggest that the compensator can improve tracking performance in the presence of packet losses, and may bring benefits to wireless control and automation applications. Further work is needed to gauge the performance of the compensator in more complex situations, for example with open-loop unstable process models and in the presence of measurable and unmeasurable disturbances. In addition, areas of future work will investigate the possibility of recursively tracking process parameters, along with improvements to the operations of the prototype WSAN platform.

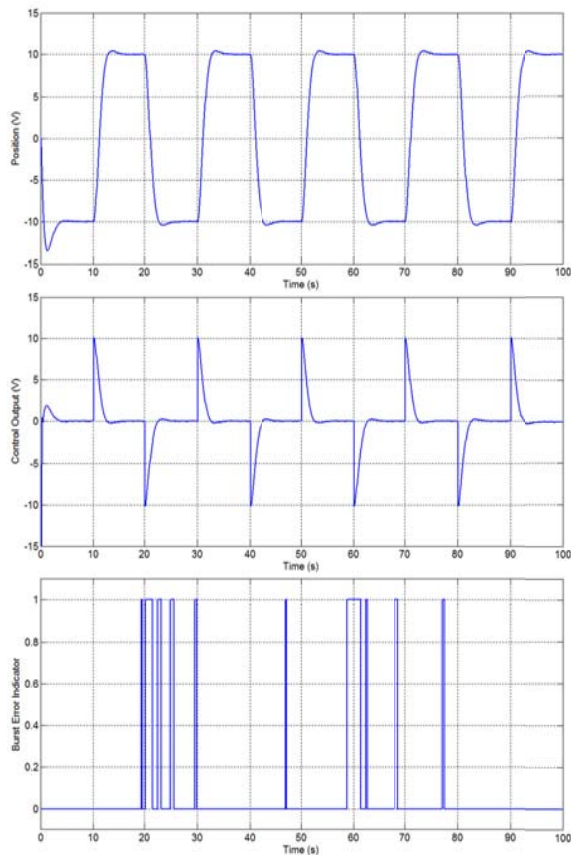


Figure 8: Effect of packet loss with compensation - refer to text for description.

#### REFERENCES

- [1] A. Flammini, P. Ferrari, D. Marioli, E. Sissini and A. Taroni. Wired and wireless sensor networks for industrial applications. *Microelectronics Journal*, Vol. 40, pp. 1322-1336, 2009.
- [2] H. A. Thompson. Wireless and Internet communications technologies for monitoring and control. *Control Engineering Practice*, Vol. 12, No. 6, pp. 781-791, 2004.
- [3] S. Munir, S. Lin, E. Hoque, S.M.S. Nirjon, J.A. Stankovic & K. Whitehouse. Addressing Burstiness for Reliable Communication and Latency Bound Generation in Wireless Sensor Networks. In: *Proceedings of the 9<sup>th</sup> ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 303-314, Stockholm, Sweden, April 2010.
- [4] W. Ikram & N.F. Thornhill. Wireless Communication in Process Automation: A Survey of Opportunities, Requirements, Concerns and Challenges. In: *Proceedings of the 8<sup>th</sup> IFAC Conference on Control (UKACC)*, Coventry, UK, September 2010.
- [5] R.S. Oliver and G. Fohler. Timeliness in Wireless Sensor Networks: Common Misconceptions. In: *Proceedings of the 9<sup>th</sup> International Workshop on Real-Time Networks (RTN)*, Brussels, Belgium, July 2010.
- [6] R. Ghostine, J.-M. Thiriet and J.-F. Aubry. Variable delays and message losses: Influence on the reliability of a control loop. *Reliability Engineering & System Safety*, Vol. 96, pp. 160-171, 2011.
- [7] S. Rothmensen. *The reality of wireless real-time: wireless networks in industrial automation*. Keynote Speech at the 9<sup>th</sup> International Workshop on Real-Time Networks (RTN), Brussels, Belgium, July 2010.
- [8] M.J. Pont. *Patterns for time-triggered embedded systems*. Addison-Wesley, 2001.
- [9] A. Albert. Comparison of event-triggered and time-triggered concepts with regard to distributed control systems. In: *Proceedings of Embedded World*, Nurnberg, Germany, 17-19 Feb, pp. 235-252.
- [10] M. Short, M.J. Pont and J. Fang. Assessment of performance and dependability in embedded control systems: methodology and case study. *Control Engineering Practice*, Vol. 16, pp. 1293-1307, 2008.
- [11] S. Gobriel, R. Cleric and D. Mosse. Adaptations of TDMA scheduling for Wireless Sensor Networks. In: *Proceedings of the 7<sup>th</sup> International Workshop on Real-Time Networks*, July 2009.
- [12] R. Costa, P. Portugal, F. Vasques and R. Moraes. A TDMA-based Mechanism for Real-Time Communication in IEEE 802.11e Networks. In: *Proceedings of ETFA'10 - 15<sup>th</sup> IEEE International Conference on Emerging Technologies and Factory Automation*, 2010.
- [13] J. Ferreira, A. Oliveira, P. Fonseca and J.A. Fonseca. An Experiment to Assess Bit Error Rate in CAN. In: *Proceedings of the 3<sup>rd</sup> International Workshop on Real-Time Networks (RTN)*, June 2004.
- [14] C.-X. Wang & W. Xu. Packet-Level Error Models for Digital Wireless Channels. In: *Proceedings of the IEEE International Conference on Communications*, pp. 2184-2189, Seoul, Korea, May 16-20, 2005.
- [15] A. Willig. A New Class of Packet- and Bit-Level Error Models for Wireless Channels. In: *Proceedings of the IEEE International Symposium on Personal, Indoor & Mobile Radio Communications*, September 2002.
- [16] K.G. Shin and H. Kim. Derivation and Application of Hard Deadlines for Real Time Control Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, pp. 1403-1413, 1992.
- [17] P.A. Chou & M. van der Schaar (eds.). *Multimedia over IP and Wireless Networks*. Elsevier Academic Press, 2007.
- [18] Y.-C. Tian & D. Levy. Compensation for control packet dropout in networked control systems. *Information Sciences*, Vol. 178, pp. 1263-1278, 2008.
- [19] H. Ishii.  $H_\infty$  control with limited communication and message losses. *Systems & Control Letters*, Vol. 57, pp. 322-331, 2006.
- [20] Q. Ling & M.D. Lemmon. Optimal Dropout Compensation in Networked Control Systems. In: *Proceedings of the 42<sup>nd</sup> IEEE Conference on Decision and Control*, pp. 670-675, Maui, Hawaii, USA, December 2003.
- [21] J.P. Norton. *An Introduction to Identification*. Academic Press, 1986.
- [22] E. Ikonen & K. Najim. *Advanced Process Identification & Control*. CRC Press, 2001.
- [23] S. Bennett. *Real-Time Computer Control: An Introduction*. Pearson Education Limited, 1988.



# Neural Generalized Predictive Controller and Internal Model Principle

Hesham Abdel-Ghaffar<sup>1</sup>, Sherif Hammad<sup>2</sup>, Hazem Abbas<sup>3</sup>, A.Z. Badr<sup>4</sup>, Ahmed Hassan<sup>5</sup>

<sup>1</sup>SCADA Group Head, <sup>2,3</sup>Mentor Graphics Egypt, <sup>4,5</sup>Systems and Control Department

<sup>1</sup>Invensys Engineering & Service, <sup>2,3</sup>Mentor Graphics Corporate, <sup>4,5</sup>Ain Shams University  
Cairo, Egypt

hesham.fouad@invensys.com, sherif\_hammad@mentor.com, hazem\_abbas@mentor.com, ahyousef@ieee.org, azbadr@ieee.org

**Abstract**—this paper derived a new formulation for Lyapunov stability analysis of Neural Generalized Predictive Controller (NGPC). Paper also applied simple technique to improve NGPC stability by using internal model of disturbance in feed forward. Finally, paper presents comparison study for the effect of using internal model principle with severe nonlinear process under different bounded disturbances. Calculation of IAE, ISE and differential Lyapunov function is used to verify stability enhancement during simulation.

**Keywords**-Neural Generalized Predictive Controller; Cost Function Minimization; Internal Model Principle; Lyapunov Stability;

## I. INTRODUCTION

The GPC, introduced by Clarke and his coworkers provides advantages over other controller types in controlling non-minimum phase plants and plants with variable or unknown dead time [1]. However, the previous work of GPC was focusing on linear process or approximated linearized process for non-linear systems. This problem has been resolved by using artificial neural network techniques in identifying complex nonlinear processes [2], [3]. Several researches on nonlinear plants proved that the ability of the GPC to make accurate predictions can be enhanced if a Neural Network (NN) is used to learn the dynamics of the plant [5], [6], [7]. Therefore there was a need to use the Nonlinear Predictor Neural Generalized Predictive Controller (NP-NGPC) method [11] or simply called NGPC.

In NGPC, another problem raised from the fact that Cost Function Minimization (CFM) algorithm used inside GPC is massive time consuming and may not be convenient of controlling fast processes. Very few papers address real-time implementation of NGPC for small time constant processes due to massive calculation problem. The Newton-Raphson algorithm [5], [9], [10] was found as one of the most efficient methods for minimizing predictive cost functions. It is quadratic algorithm converging better than others. It requires less iteration numbers, usually two iterations only, for convergence and this reduces the calculation time.

Although well mathematical structure of NGPC was presented in previous papers, the stability analysis under output disturbances was not covered. Actually there are few papers studied the Lyapunov stability of GPC using recurrent neural network and Adaptive Learning Rate

(ALR) [12]. Some others studied the Lyapunov stability of predictive controllers using either receding horizon controllers [17], [18] or neural controllers [4], [19]. However none of the above papers covered stability analysis of neural GPC using feed forward neural network.

In this paper, new formulation for Lyapunov Stability analysis of NGPC controlling nonlinear process is derived. Also Internal Model Principle (IMP) technique [13], [14], [15] is applied with NGPC in order to enhance disturbance mitigation. Finally simulation analysis was used to verify Lyapunov conditions derived and IMP enhancements. The rest of this paper is organized as follows. Section II shall calculate new Lyapunov stability conditions for neural GPC controlling general nonlinear process. Section III presents comparison simulation analysis for using IMP with NGPC. Section IV concludes the results and limitations.

## II. LYAPUNOV STABILITY ANALYSIS

The NGPC system starts with the input signal,  $r(n)$ , which is presented to the reference model. This model produces a tracking reference signal,  $y_m(n)$  that is used as an input to the Cost Function Minimization (CFM) block. The CFM used in this paper is Newton-Raphson numerical minimization algorithm. The CFM produces an output that is either used as an input to the plant or the plant's neural model. The double pole double throw switch,  $S$ , is set to the plant when the CFM algorithm has solved for the best control input,  $u(n)$ , that will minimize a specified cost function. Between samples, the switch is set to the plant neural model where the CFM algorithm uses this model to calculate the next control input,  $u(n+1)$ , from predictions of the response of neural model  $y_n(n)$ . Once the cost function is minimized, this input is passed to the plant. The NGPC cost function can be represented by:

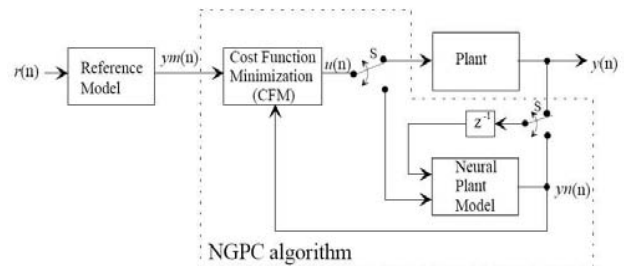


Fig. 1: Neural GPC structure block diagram.

$$J(N_1, N_2, N_u) = \delta \sum_{j=N_1}^{N_2} (y_d(k+j) - \hat{y}(k+j))^2 + \lambda \sum_{j=1}^{N_u} (\Delta u(k+j))^2 \quad (1)$$

$N_1$  = Minimum Prediction Horizon ( $N_1=d+1$ ).

$d$ = Open loop Input/output process delay. Usually  $d=0$ .

$N_2$  = Maximum Prediction Horizon.

$N_u$ = Maximum Control Horizon.

$\hat{y}(k)$  = Predicted Output from Neural Network at sample time ( $k$ ). This is equivalent to  $y_n$  in Fig. 1.

$u(k)$  = Manipulated Control Input at sample time ( $k$ ).

$\Delta u(k)$  = variation in control input at sample time ( $k$ ) =  $u(k+1) - u(k)$ .

$y_d(k)$  = Reference Trajectory at sample time ( $k$ ). This is equivalent to  $y_m$  in Fig. 1.

$\delta$  and  $\lambda$  = Weighing Factor for tracking error and control input respectively. For simplicity we consider  $\lambda$  constant across all sample times and  $\delta = 1$ .

Assuming discrete nonlinear state space equations:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k));$$

$$\mathbf{y}_p(k) = \mathbf{h}(\mathbf{x}(k), \mathbf{u}(k)) \quad (2)$$

Let us assume SISO NGPC as neural MPC controller has admissible control input  $u(k)$ . The admissible term here for  $u(k)$  means that control input should verify stability according to Lyapunov theorem as well as minimizing predictive cost function of NGPC in (1).

Putting (1) in vector notation lead to:

$$J(k) = [\mathbf{Y}_d(k) - \mathbf{Y}(k)]^T * [\mathbf{Y}_d(k) - \mathbf{Y}(k)] + \lambda \Delta \mathbf{U}(k)^T * \Delta \mathbf{U}(k) \quad (3)$$

Where:

- Reference trajectory at instant ( $k$ ),  
 $\mathbf{Y}_d(k) = [y_d(k+1) \ y_d(k+2) \dots y_d(k+N_2)]^T$
- Predicted horizon NN output at instant ( $k$ ),  
 $\mathbf{Y}(k) = [\hat{y}(k+1) \ \hat{y}(k+2) \dots \hat{y}(k+N_2)]^T$
- Control trajectory variation at instant ( $k$ ),  
 $\Delta \mathbf{U}(k) = [\Delta u(k+1) \ \Delta u(k+2) \dots \Delta u(k+N_u)]^T$
- Predicted control Input at instant ( $k$ ),  
 $\mathbf{U}(k) = [u(k+1) \ u(k+2) \dots u(k+N_u)]^T$
- $\Delta u(k+1) = [u(k+1) - u(k)]$  and  
 $\Delta u(k+N_u) = [u(k+N_u) - u(k+N_u-1)]$
- Control weighing factor  $\lambda \geq 0$ .  $N_1=1$  and  $N_2 \geq N_u$ .

Also let us assume the Lyapunov candidate function as:

$$L(k) = [\mathbf{Y}_d(k) - \mathbf{Y}(k)]^T * [\mathbf{Y}_d(k) - \mathbf{Y}(k)]$$

$$\mathbf{L}(k) = \mathbf{E}^T(k) * \mathbf{E}(k) \quad (4)$$

Where  $\mathbf{E}(k) = [e(k+1) \ e(k+2) \dots e(k+N_2)]$  and

$$e(k+1) = y_d(k+1) - \hat{y}(k+1).$$

It is apparent that  $L(k) > 0$  positive definite which is first necessary condition for Lyapunov stability of closed loop NGPC. The second necessary condition is  $\Delta L(k) \leq 0$ .

$$\Delta \mathbf{L}(k) = \mathbf{L}(k+1) - \mathbf{L}(k) = 2\Delta \mathbf{E}^T(k) * \mathbf{E}(k) + \Delta \mathbf{E}^T(k) * \Delta \mathbf{E}(k) \quad (5)$$

Where  $\mathbf{E}(k) = [\mathbf{Y}_d(k) - \mathbf{Y}]$  ;  $\mathbf{E}(k+1) = \mathbf{E}(k) + \Delta \mathbf{E}$ , and,

$$\Delta \mathbf{E}(k) = \frac{\partial \mathbf{E}(k)}{\partial \mathbf{U}(k)} * \Delta \mathbf{U}(k) = -\frac{\partial \mathbf{Y}(k)}{\partial \mathbf{U}(k)} * \Delta \mathbf{U}(k) = -\mathbf{G}(k) * \Delta \mathbf{U}(k) \quad (6)$$

Where  $\frac{\partial \mathbf{Y}(k)}{\partial \mathbf{U}(k)} = \mathbf{G}(k)$  represent controller gain matrix and

has size  $[N_2 \times N_u]$ . From matrix algebra:

$$\mathbf{G}(k) = \begin{pmatrix} \frac{\partial [\hat{y}(k+1)]}{\partial u(k+1)} & \dots & \frac{\partial [\hat{y}(k+1)]}{\partial u(k+N_u)} \\ \vdots & \ddots & \vdots \\ \frac{\partial [\hat{y}(k+N_2)]}{\partial u(k+1)} & \dots & \frac{\partial [\hat{y}(k+N_2)]}{\partial u(k+N_u)} \end{pmatrix} \quad (7)$$

According to Newton-Raphson CFM algorithm [9]

$$\Delta \mathbf{U}(k) = -\eta * [\mathbf{H}]^{-1} * \frac{\partial J(k)}{\partial \mathbf{U}(k)} = -\eta * \mathbf{H} \text{inv} * \frac{\partial J(k)}{\partial \mathbf{U}(k)} \quad (8)$$

Where:

- $\partial J/\partial \mathbf{U}$ : is the cost function partial differential with respect to control vector denoted by Jacobian and has size of  $[N_u \times 1]$ .
- $\mathbf{H}$ : is the cost function second partial differential with respect to control vector denoted by Hessian. The Hessian matrix is always symmetric and has size of  $[N_u \times N_u]$ .
- $\mathbf{H} \text{inv}$ : is the inverse Hessian and also symmetric.
- $\eta$ : Cost function line search optimization step parameter.

The cost function Jacobian can be expressed as:

$$\frac{\partial J(k)}{\partial \mathbf{U}(k)} = -2 * \left[ \frac{\partial \mathbf{Y}(k)}{\partial \mathbf{U}(k)} \right]^T * \mathbf{E}(k) + 2 * \lambda * \frac{\partial \Delta \mathbf{U}(k)}{\partial \mathbf{U}(k)} * \Delta \mathbf{U}(k) \quad (9)$$

$$\frac{\partial \Delta \mathbf{U}(k)}{\partial \mathbf{U}(k)} = \begin{pmatrix} \frac{\partial [u(k+1) - u(k)]}{\partial u(k+1)} & \dots & \frac{\partial [u(k+N_u) - u(k+N_u-1)]}{\partial u(k+1)} \\ \vdots & \ddots & \vdots \\ \frac{\partial [u(k+1) - u(k)]}{\partial u(k+N_u)} & \dots & \frac{\partial [u(k+N_u) - u(k+N_u-1)]}{\partial u(k+N_u)} \end{pmatrix} \quad (10)$$

As  $\frac{\partial \Delta u(k+j)}{\partial u(k+h)} = \delta(h, j) - \delta(h, j-1)$ , Then

$$\frac{\partial J(k)}{\partial \mathbf{U}(k)} = -2 * \mathbf{G}^T(k) * \mathbf{E}(k) + 2 * \lambda * \mathbf{A} * \Delta \mathbf{U}(k) \quad (11)$$

Where  $A$ : is constant square matrix ( $N \times N$ ) independent of sample time  $k$ . Substituting Jacobian above in (8):

$$\Delta U(k) = 2\eta * Hinv * G^T(k) * E(k) - 2\eta\lambda * Hinv * A * \Delta U(k) \quad (12)$$

$$\Delta U(k) = 2\eta * V * Hinv * G^T(k) * E(k) \quad (13)$$

Where  $V = [I + 2\eta\lambda * Hinv * A]^{-1}$  is square non-symmetric matrix with size  $[N \times N]$ .

Substituting  $\Delta U(k)$  in (6):

$$\Delta E(k) = -2\eta * G(k) * V * Hinv * G^T(k) * E(k) \quad (14)$$

$$\Delta E^T(k) = -2\eta * E^T(k) * G(k) * H^T inv * V^T * G^T(k) \quad (15)$$

Substituting  $\Delta E(k)$  &  $\Delta E^T(k)$  in (5) we can find:

$$\begin{aligned} \Delta L &= -4\eta * E^T(k) * G(k) * H^T inv * V^T * G^T(k) * \\ &E(k) + 4\eta^2 \{ E^T(k) * G(k) * H^T inv * V^T * G^T(k) * \\ &G(k) * V * Hinv * G^T(k) * E(k) \} \\ \Delta L &= -4\eta * E^T(k) * [G(k) * H^T inv * V^T * G^T(k) * \\ &[I - \eta * G(k) * V * Hinv * G^T(k)]] * E(k) \end{aligned} \quad (16)$$

From matrix algebra if matrix  $P$  is symmetric positive definite then  $A^T * P * A$  is positive function for all vectors  $A \neq 0$ . Also all eigenvalues of matrix  $P$  must be positive.

Therefore to obtain  $\Delta L \leq 0$  and verify Lyapunov stability the following positive condition should verify:

$$\{G(k) * H^T inv * V^T * G^T(k) - \eta * G(k) * H^T inv * V^T * G^T(k) * G(k) * V * Hinv * G^T(k)\} \geq 0 \quad (17)$$

The difficulty of above proof come from the fact that matrix  $V$  is not always symmetric. This is because, although  $Hinv$  matrix is always symmetric, the matrix ( $Hinv * A$ ) is non-symmetric and so matrix  $V$  generally non-symmetric. However there is a fact that in some cases where  $\eta$  is very small,  $V$  can be approximated to unity matrix.

**Assumption 1:** For the simplicity of calculation and not affecting the flow of derivation, when  $\eta$  is very small  $\Delta U(k)$  is very small and  $\{A * \Delta U(k)\} \approx \Delta U(k)$ . Then

$$V \approx Hinv \approx I_{N \times N} \quad (18)$$

And Lyapunov equality in (17) can be simplified to:

$$\{G(k) * G^T(k) - \eta * G(k) * G^T(k) * G(k) * G^T(k)\} \geq 0 \quad (19)$$

As left hand side of equality is always symmetric matrix, therefore to achieve positive definiteness of this matrix, the eigenvalues must be all positive.

$$\{eig(G(k) * G^T(k)) - \eta * eig(G(k) * G^T(k) * G(k) * G^T(k))\} \geq 0 \quad (20)$$

Where  $eig(A)$  is eigenvalues of matrix  $A$ . From matrix algebra this requires:

$$\eta < \frac{1}{\sigma_{\max}^2}, \forall k > k_0 \quad (21)$$

Where  $\sigma_{\max}$  is the maximum singular value decomposition of non square controller gain matrix  $G(k)$ . The above approximation is the same as (Chi-Huang Lu) obtained in normal gradient descent algorithm using recurrent neural identification network model [12].

Therefore in case of validity approximation made in (18), the second Lyapunov stability condition ( $\Delta L \leq 0$ ) is valid if and only if NGPC optimum line search step “ $\eta$ ” at each sample time verifies (21). This means to get Asymptotic Stability (AS) or Stable In Sense of Lyapunov (SISL) at any sample time ( $k$ ), maximum singular value of controller gain matrix should verify:

$$\sigma_{\max}(k) < \frac{1}{\sqrt{\eta(k)}} \quad (22)$$

In case we add constraint  $E(k) > 0$ ; This leads to the same Lyapunov theorem in [4] for Uniform Ultimately Bounded (UUB) Stability.

*The mathematic challenge is raised when  $\eta$  is not very small and hence approximation made is no longer valid and sophisticated calculation of condition mentioned in (17) is required at each sample time to verify stability.*

### III. SIMULATION RESULTS

The nonlinear process under study is severe with non-minimum phase and variable dead time.

$$\ddot{y} + \dot{y} + y + y^3 = 2u - \dot{u} \quad (23)$$

The target from this section is to verify by simulation the stability derived conditions of (17) and (22). Also the simulation results showed the substantial improvement of disturbance mitigation by using IMP technique with NGPC. Simulation proved that Lyapunov stability approaches UUB criteria under bounded disturbance. The parameters used in simulation mentioned in table below.

Table I: NGPC Simulation Parameters.

NGPC Parameters	Simulation value
Sampling Time ( $T_s$ )	0.1s
Prediction Horizon ( $N_2$ )	50
Control Horizon ( $N_u$ )	10
Control Weighing ( $\lambda$ )	0.3
Search Parameter ( $\alpha$ )	0.001
Iterations per sample	2
Hidden Layer Neurons ( $L$ )	9
Delayed Plant Output ( $N_j$ )	4
Delayed Control Input ( $N_i$ )	4
Control Input Allowed	$-4 \leq u(t) \leq +4$
Disturbance Filter	$16/(s^2 + 5.6s + 32)$
Signal/Noise ratio	17-23 dB
Training samples	4,000
Training Epochs	1,000

Minimization Routine	Newton-Raphson
NN Training Algorithm	Levenberg-Marquardt
NN Modeling Error ( $\epsilon$ )	$\leq 3e-04$

As seen from Fig. 2, Fig. 5 and Fig. 8 the applied

reference is step input,  $y_d(t) = \begin{cases} 0.5, & t < 75 \\ 1, & t \geq 75 \end{cases}$

#### A. NGPC Stability without Disturbance

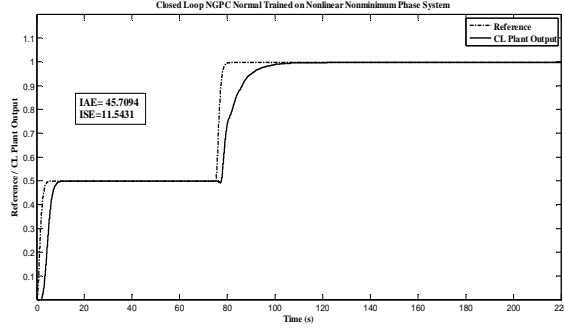


Fig. 2: Closed loop NGPC **without** disturbance.

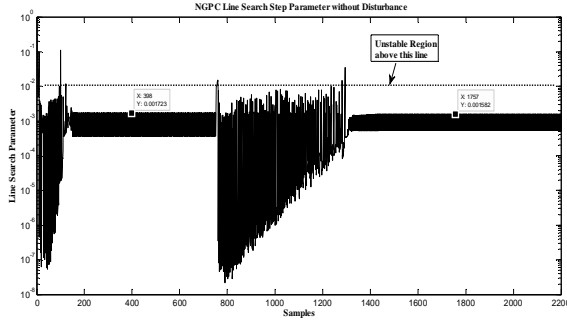


Fig. 3: Line search step ( $\eta$ ) **without** disturbance.

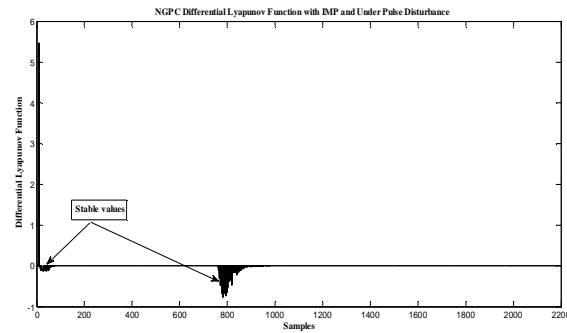


Fig. 4: Lyapunov ( $\Delta L$ ) **without** disturbance.

As seen from above figures, NGPC is able to stabilize this severe nonlinear system, non-minimum phase and variable dead time, keeping ( $\eta$ ) limited in stable range and so maintain  $\Delta L(k) \leq 0$  for all  $k > k_0$ . As the approximation made in (18) is applicable in above case, the dotted line in Fig. 3 represent stable margin  $1/\sigma_{\max}^2$ . Therefore the values of ( $\eta$ ) above this margin make the system unstable.

#### B. NGPC Stability with disturbance and without IMP

The disturbance is pulse (Period=40 s, Amp= 0.14, Pulse width= 12s) applied on nonlinear plant output. The disturbance filter is dividing the amplitude of disturbance pulse by 2. Therefore applied disturbance amplitude is actually 0.07. This is to get S/N ratio around 20 dB (i.e. disturbance  $\leq 10\%$  Reference).

In Fig. 5, NGPC normal trained on nonlinear process without training on disturbance dynamics or including IMP technique is not able to reject the disturbance impact on plant output.

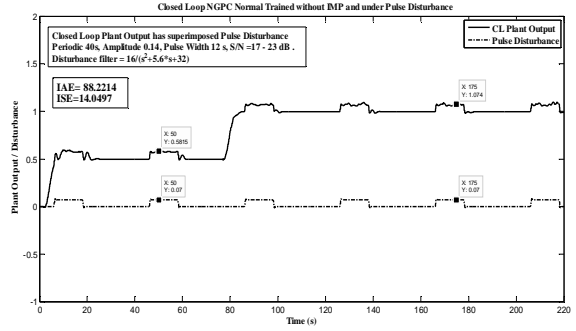


Fig. 5: Closed loop NGPC **without** IMP and under pulse disturbance.

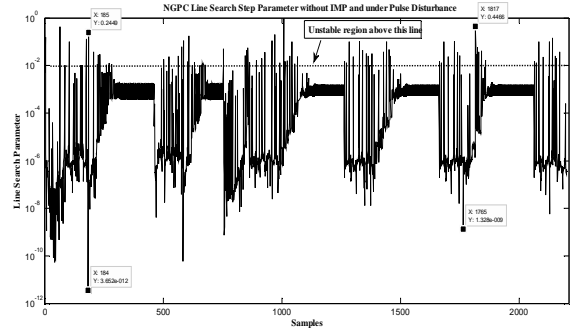


Fig. 6: Line search step ( $\eta$ ) **without** IMP and under pulse disturbance.

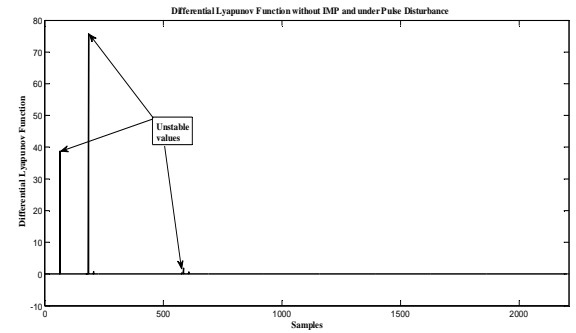


Fig. 7: Lyapunov ( $\Delta L$ ) **without** IMP and under pulse disturbance.

As seen, the same (0.07) disturbance value applied appear in plant output response accompanied with bad dynamics of disturbance filter. In Fig. 6 the line search parameter ( $\eta$ ) is bumping very fast exceeding stability range of NGPC during pulse disturbance. In Fig. 7 system is not stable as  $\Delta L(k) \geq 0$  especially when S/N ratio is less than 20dB during step reference (0.5).

### C. NGPC Stability with disturbance and with IMP

In Fig. 8 NGPC is trained on both nonlinear process and disturbance filter including IMP technique.

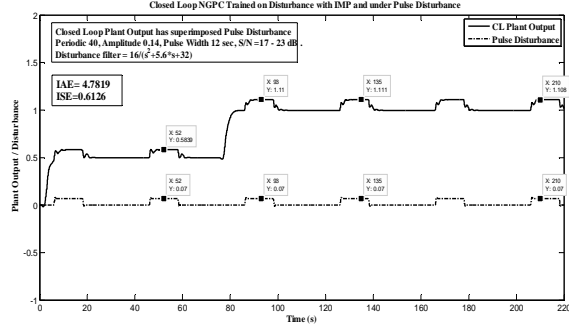


Fig. 8: Closed loop NGPC with IMP and under pulse disturbance.

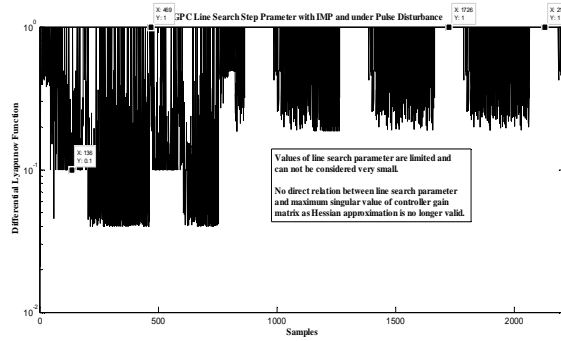


Fig. 9: Line search step ( $\eta$ ) with IMP and under pulse disturbance.

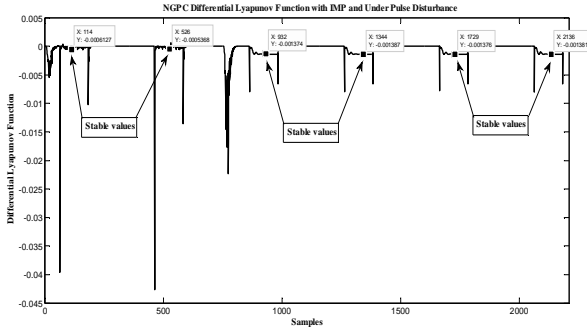


Fig. 10: Lyapunov ( $\Delta L$ ) with IMP and under pulse disturbance.

It is apparently clear from Fig. 8 and Fig. 9 that NGPC removes bad disturbance dynamics and improve CFM line search parameter ( $\eta$ ). Although this improve overall tracking and achieve Lyapunov stability, still NGPC output has some bounded tracking error due to applied pulse disturbance. Fig. 10 show that  $\Delta L < 0$  almost all the time even with low S/N ratio. This achieves Uniform Ultimately Bounded Lyapunov stability under bounded disturbance if  $E(k) > 0$ .

The input/output training samples for NGPC, including disturbance internal model, are obtained from Fig. 11. The disturbance dynamics selected for example in this simulation is second order stable filter. To obtain good results, the functional reconstruction error,  $\epsilon(k)$ , of NGPC shall be as minimum as possible for all  $k > k_0$  as seen in Fig. 12. Where:  $\epsilon(k) = y_p(k) - \hat{y}(k)$ .

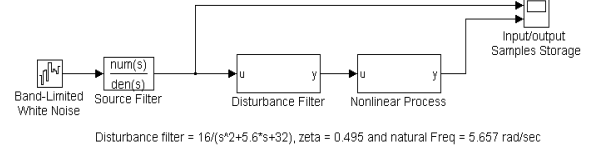


Fig. 11: Training Samples of nonlinear process including IMP.

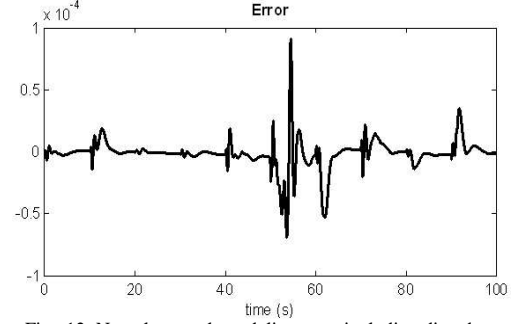


Fig. 12: Neural network modeling error including disturbance filter.

Table II offers comparative simulation study for Lyapunov stability for the same nonlinear system under pulse and sinusoidal disturbances.

Table II: NGPC stability performance under variable disturbances.

NGPC Stability Performance for NonLinear Process under study					
	Step Ref	( $\eta$ )	Average $1/\sigma^2_{max}$	NGPC Stability & Performance Results	
Without disturbance	Ref (0.5)	$3.625e-04 \leq \eta \leq 1.724e-03$	0.0152	negative AL tends to zero	IAE=45.7094 ISE=11.5431
	Ref (1)	$5.342e-04 \leq \eta \leq 1.582e-03$	0.0129	negative AL tends to zero	As $\eta < 1$ , $V \approx H_{inv} \approx 1$ , and to get stability $\eta < 1/\sigma^2_{max}$ should verify
Slow periodic pulse disturbance and No IMP	Ref (0.5)	$3.652e-012 \leq \eta \leq 0.2449$ during disturbance, elsewhere same as without disturbance	0.0146	positive AL tends to zero, unstable behavior	IAE=88.2214 ISE=14.0497
	Ref (1)	$1.328e-09 \leq \eta \leq 0.4466$ during disturbance, elsewhere same as without disturbance	0.0129	AL tends to zero, boundary of instability	As $\eta < 1$ , $V \approx H_{inv} \approx 1$ , and to get stability $\eta < 1/\sigma^2_{max}$ should verify
Medium / fast (4 rad/sec) periodic sin disturbance and No IMP ( $\lambda=0.3$ , $N2=50$ , $Nu=10$ , $\alpha=0.001$ )	Ref (0.5)	$1.6e-13 \leq \eta \leq 0.4236$ during disturbance	0.0114	positive AL tends to zero, unstable behavior	IAE=181.3263 ISE=24.9793
	Ref (1)	$7.147e-14 \leq \eta \leq 0.5$ during disturbance	0.0130	positive AL tends to zero, unstable behavior	As $\eta < 1$ , $V \approx H_{inv} \approx 1$ , and to get stability $\eta < 1/\sigma^2_{max}$ should verify
Slow periodic pulse disturbance and with IMP plus disturbance training	Ref (0.5)	$(0.0209 \leq \eta \leq 1)$ generally and $0.1 \leq \eta \leq 1$ during disturbance	Not Needed, approx. is not valid and no direct relation between $\eta$ & $\sigma^2_{max}$	negative AL tends to -0.0001329 during disturbance and zero without disturbance	IAE=4.7819 ISE=0.6126
	Ref (1)	$(0.1868 \leq \eta \leq 1)$ generally and $\eta=1$ during disturbance		negative AL tends to -0.001393 during disturbance and zero without disturbance	Improve tracking stability using IMP. As $V \neq H_{inv} \neq 1$ , therefore to get stability $\Delta L$ should be negative

NGPC Stability Performance for NonLinear Process under study					
	Step Ref	( $\eta$ )	Average $1/\sigma_{\max}^2$	NGPC Stability & Performance Results	
Medium sin (4rad/sec) disturbance and with IMP plus disturbance training ( $\lambda=1$ , $N_2=30$ , $N_u=10$ , $\alpha=0.001$ )	Ref (0.5)	$0.2321 \leq \eta \leq 1$	Not Needed, approx. is not valid and no direct relation between $\eta$ & $\sigma_{\max}^2$	negative $\Delta L$ tends to zero	IAE=3.6569 ISE=0.4949
	Ref (1)	$0.2411 \leq \eta \leq 1$		negative $\Delta L$ tends to zero	Improve tracking stability using IMP. As $V \neq Hinv \neq I$ , therefore to get stability $\Delta L$ should be negative

#### IV. CONCLUSION

To have successful NGPC control, Lyapunov stability conditions derived in this paper should be satisfied. Normal trained NGPC may not be able alone to stabilize severe nonlinear process under output disturbance. Applying IMP improves stability of both tracking and bounded disturbance mitigation. To have successful implementation of IMP some conditions should be achieved: (1) Closed loop NGPC for nonlinear process without disturbance should be stable. (2) Output disturbance is bounded and disturbance dynamics are known and stable. (3) Desired trajectory reference is bounded. (4) NGPC neural network should be trained on both nonlinear process and disturbance dynamics. This means that function reconstruction error or neural network modeling error including disturbance filter should be bounded and very small.

The paper proved that in some cases where line search parameter ( $\eta$ ) is very small, the stability criteria depend only on maximum singular value ( $\sigma_{\max}$ ) of controller gain matrix.

#### REFERENCES

- [1] D. W. Clarke, C. Mohtadi and P. S. Tuffs: "Generalized predictive control-Parts I & II. The basic algorithm", Automatica, vol 23, pp. 137-163. 1987.
- [2] R K Al Seyab; Y Cao: "Nonlinear system identification for predictive control using continuous time recurrent neural networks and automatic differentiation", Journal of Process Control, Volume 18, Issue 6, July 2008, Pages 568-581.
- [3] David T. Westwick and Robert E. Kearney: "Identification of nonlinear physiological systems", Institute of Electrical and Electronics Engineers, 2003, chapters 6, 7 & 8.
- [4] Jagannathan Sarangapani: "Neural network control of nonlinear discrete-time systems", University of Missouri, Rolla, Missouri, chapters 1, 3, 4 & 7.
- [5] Sadhana CHIDRAWAR, Balasaheb PATRE: "Generalized predictive control and neural generalized predictive control", Leonardo Journal of Sciences, Issue 13, July-December 2008, Pages 133-152.
- [6] S. Chen, S.Billings, P. Grant: "Non-linear system identification using neural networks", International Journal of Control 51, No. 6, pp. 1191-1214, 1990.
- [7] K. S. Narendra, A. U. Levin: "Identification and control of dynamical systems using neural networks", IEEE Transaction on Neural Networks 1, No. 1, pp. 4-27, 1990.
- [8] Anna Vasičkaninová, Monika Bakošová: "Neural network predictive control of a chemical reactor", Institute of Information Engineering, Automation and Mathematics,

Faculty of Chemical and Food Technology, Acta Chimica Slovaca, Vol.2, No.2, 2009, pp 21 – 36.

- [9] Donald Soloway & Pamela J. Haley: "Neural generalized predictive control: A Newton-Raphson implementation", NASA Technical Memorandum 110244, Feb 1997.
- [10] B. Durmuş, H. Temurtaş, N. Yumuşak, F. Temurtaş, R. Kazan: "The Cost Function Minimization for predictive control by Newton-Raphson method", Proceedings of IMECS 2008, Vol. II.
- [11] D.N.Rao, M.R.K.Murthy, S.R.M.Rao, D.N.Harshal: "Comparison of NGPC with approximate and nonlinear predictive control - A simulation study", ACSE Journal, Volume (6), Issue (1), January, 2006.
- [12] Chi-Huang Lu, "Discrete time neural predictive controller design", HSIUPING JOURNAL. Volume (18), pp 27-38, March 2009.
- [13] Jie Huang, "Internal models for nonlinear systems: an overview", ICCM 2007, Vol. III, 669–681.
- [14] Jie Huang and Ching-Fang Lin, "Internal model principle and robust control of nonlinear Systems", Proceedings of 32nd Conference on Decision and Control, San Antonio, Texas, December 1993.
- [15] Jie Huang, "Internal model principle: from output regulation to stabilization", International Conference on Control, Automation and Systems October 2007, Seoul, Korea.
- [16] MathWorks, "MATLAB 7.8 user guide", Neural Network Toolbox, Neural Predictive Controller, Release R2009a.
- [17] Michalska, H. Mayne, D.Q., "Robust receding horizon control of constrained nonlinear systems", IEEE Transactions on automatic control, No. 11, 38, 1993.
- [18] Jim Benjamin Luther, "Stability of a Neural Predictive Controller Scheme on a Neural Model", Department of automation (IAU), Technical university of Denmark (DTU), 1999.
- [19] Shuzhi Sam Ge, Jin Zhang, and Tong Heng Lee "Adaptive Neural Network Control for a Class of MIMO Nonlinear Systems With Disturbances in Discrete-Time", IEEE, Part B: Cybernetics, Vol. 34, No. 4, August 2004.
- [20] Dennis and Schnabel, "Numerical Methods for Unconstrained Optimization and Nonlinear Equations", 1983.



# State Observer in Networked Control Systems with Variable Delay in the Feedback Channel

A. Sedighi, R. Mahboobi Esfanjani

Department of Electrical Engineering

Sahand University of Technology

Tabriz, Iran

[A\\_Sedighi@sut.ac.ir](mailto:A_Sedighi@sut.ac.ir), [mahboobi@sut.ac.ir](mailto:mahboobi@sut.ac.ir)

**Abstract-** This paper considers the design of state observer for discrete linear time invariant system where sensor measurements are transmitted via a communication network. Both of the data packet delay and dropout are considered in the Networked Control System (NCS) model. The dynamics of estimator is modeled in the switched systems' framework. The issue of estimator design is regarded as the stabilization problem of a switched system. Estimator gain is determined by solving Bilinear Matrix Inequality (BMI). A numerical example is utilized to demonstrate the effectiveness and advantages of this scheme.

**Keywords-** Networked Control System (NCS); State Observer; Variable Network Delay; Switched System.

## I. INTRODUCTION

A Networked Control System (NCS) is a feedback control system in which controller and spatially distributed sensors and actuators exchange information via a communication network. The advantages of NCS such as low cost, simple installation and maintenance make it more and more popular in many applications including manufacturing plants, intelligent traffic systems, cluster of unmanned air vehicles and other multi-agent systems. However, the presence of communication network in control loop make the analysis and design of the NCS more complicated than classical control systems. Main issues in this paradigm are the packet dropouts and delays in the communication channels [1], [2].

Various approaches have been developed in the recent years to design of observer for NCS [3]-[9]. In [3], networked control system was designed for continuous-time systems, where the measurement channel is subject to random sensor delay. A design scheme for the observer-based output feedback controller is proposed to render the closed-loop networked system stable which is based on solving linear matrix inequalities. In [4], an observer-protocol pair was designed to asymptotically reconstruct the states of a linear time-invariant plant under communication constraints induced by the network. Sufficient conditions were derived in terms of matrix inequalities for the existence of an observer-protocol pair in the considered class. In [5], optimal estimator design was studied for sampled linear systems where sensor

measurements are subject to random delay or might even be completely lost. Alternative estimator architectures were presented which are computationally efficient. In [6], the principle of predictive control is adopted to overcome the effects of random network time delay in the feedback channel. Moreover, it was shown that the closed-loop networked predictive control system with bounded random network delay is stable if the corresponding switched system is stable. In [7]-[9] robust observers were developed for some classes of nonlinear discrete systems with missing measurements.

In this note, a networked observer is designed for discrete linear time invariant system, where the measurements are subject to transmission delay and loss. The proposed scheme is based on Luenberger observer and estimation error dynamics is formulated as a switched linear systems. Thereafter, the observer gain is determined utilizing the stability theory of switched systems. Sufficient condition is derived in terms of matrix inequality to compute the observer gain. A numerical example is used to verify the effectiveness and the merits of the presented result.

The paper is organized as follows: In section II, the system model and networked observer is described. Section III presents the main results of this paper wherein the observer gain is determined using the facts of switched system theory. In section IV, the suggested design method is illustrated by a numerical example. Finally, section V concludes this paper.

## II. PROBLEM FORMULATION

Consider the discrete linear time invariant system as follows:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t \end{aligned} \quad (1)$$

where  $x_t \in R^n$ ,  $u_t \in R^m$ , and  $y_t \in R^l$  are the state, input, and output vectors of system, respectively.  $A \in R^{n \times n}$ ,  $B \in R^{n \times m}$  and  $C \in R^{l \times n}$  are constant matrices. Moreover, the pair  $(A, C)$  is observable.

The plant (1) is equipped with a sensor which transmits the measurements through a communication network to the observer. In the considered network, all the data packets are time-stamped. Variable delay is assumed for

network induced delay to model both the data packet latency and lost. Moreover, the upper bound of the network delay is not greater than  $N$ , where  $N$  is a known positive integer.

**Remark 1:** In a Networked Control System, if the data packet does not arrive with a certain transmission delay, it means this data packet is lost. Furthermore, measurement output  $y_{t-j}$  will be used if  $y_{t-i}$  arrives after  $y_{t-j}$ , for  $j < i$ .

The Luenberger state observer is considered as follows:

$$\hat{x}_{t+1|t} = A \hat{x}_{t|t-1} + B u_t + L (y_t - C \hat{x}_{t|t-1}) \quad (2)$$

where  $\hat{x}_{t+1|t} \in R^n$  is the one-step-ahead state prediction and  $u_t \in R^m$  is the observer input. In what follows, a structured approach will be proposed to design the observer gain matrix  $L \in R^{n \times l}$ .

Motivated by [6], regarding the state observer (2) and based on the output data received at  $t - k$ , the state predictions from time  $t - k + 1$  to  $t$  is as follows:

$$\begin{aligned} \hat{x}_{t-k+1|t-k} &= A \hat{x}_{t-k|t-k-1} + B u_{t-k} + L (y_{t-k} - C \hat{x}_{t-k|t-k-1}) \\ \hat{x}_{t-k+2|t-k} &= A \hat{x}_{t-k+1|t-k} + B u_{t-k-1} \\ &\vdots \\ \hat{x}_t|t-k &= A \hat{x}_{t-1|t-k} + B u_{t-1} \end{aligned} \quad (3)$$

which results in the following relation:

$$\begin{aligned} \hat{x}_{t-k+i|t-k} &= A^{i-1}(A - LC)\hat{x}_{t-k|t-k-1} + A^{i-1}L y_{t-k} \\ &\quad + \sum_{j=1}^i A^{i-j} B u_{t-k+j-1} \end{aligned} \quad (4)$$

for  $i = 1, 2, 3, \dots, k$ .

Based on (1) and (3), estimation error is defined at sampling instants  $t - i$ , for  $i = -1, 0, 1, \dots, k$  as follows:

$$\begin{aligned} e_{t-k} &= x_{t-k} - \hat{x}_{t-k|t-k-1}, \\ e_{t-k+1} &= x_{t-k+1} - \hat{x}_{t-k+1|t-k} = (A - LC)e_{t-k} \\ &\vdots \\ e_t &= x_t - \hat{x}_t|t-k = A^{k-1}(A - LC)e_{t-k} \\ e_{t+1} &= x_{t+1} - \hat{x}_{t+1|t-k} = A^k(A - LC)e_{t-k} \end{aligned} \quad (5)$$

Now, let  $\varepsilon(t)$  be as follows:

$$\varepsilon(t) = [e_t^T \ e_{t-1}^T \ \dots \ e_{t-k}^T]^T. \quad (6)$$

Combining the relations in (5) yields to:

$$\varepsilon(t+1) = \Lambda \varepsilon(t) \quad (7)$$

where,

$$\Lambda = \begin{bmatrix} 0 & \dots & 0 & A^k(A - LC) \\ 0 & \dots & 0 & A^{k-1}(A - LC) \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & (A - LC) \end{bmatrix}$$

### III. MAIN RESULT

When the network delay in the feedback channel is constant, the estimation error dynamics  $\varepsilon(t)$  will end to zero if all the eigenvalues of matrix  $\Lambda$  are in the unit circle. But in more realistic problem where network delay is variable, the delay is represented by  $k_t \in \{0, 1, 2, \dots, N\}$ , where  $N$  is assumed to be the upper bound of variable delay. To obtain error dynamics in the case of variable delay, the error vector is defined as follows:

$$\Xi(t) = [e_t^T \ e_{t-1}^T \ \dots \ e_{t-k_t}^T \ \dots \ e_{t-N}^T]^T. \quad (8)$$

So, regarding (7), the following is obtained:

$$\Xi(t+1) = \Lambda(k_t) \Xi(t) \quad (9)$$

where  $\Lambda(k_t) \in R^{n(N+1) \times n(N+1)}$  and

$$\Lambda(k_t) = \begin{bmatrix} 0 & \dots & 0 & A^{k_t}(A - LC) & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & A^{k_t-1}(A - LC) & 0 & \dots & 0 & 0 \\ \vdots & & & \vdots & & & \vdots & \\ 0 & \dots & 0 & (A - LC) & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & I & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & I & \dots & 0 & 0 \\ \vdots & & & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & I & 0 \end{bmatrix}$$

for  $k_t \in \{0, 1, 2, \dots, N\}$ .

The autonomous discrete linear time-invariant (LTI) switched system is a collection of discrete-time LTI systems and given by

$$X(t+1) = \mathcal{A}_i X(t), \quad i \in \mathfrak{I} \quad (10)$$

where index set  $\mathfrak{I}$  is a finite set of integers and stands for the collection of discrete modes [10]. The equivalence of (9) and (10) is obvious by making  $\mathcal{A}_1 = \Lambda(k_t = 0)$ ,  $\dots$ ,  $\mathcal{A}_i = \Lambda(k_t = i - 1)$ ,  $\dots$ ,  $\mathcal{A}_{N+1} = \Lambda(k_t = N)$ . So, the dynamic equation (9) can be interpreted as a switched system consists of  $N + 1$  subsystems with arbitrary switching between these subsystems. Therefore, the theory of switched systems' stability can be used to determine observer gain matrix  $L$ .

A remarkable fact about switched systems is that even when all the subsystems are stable, the switched system may have divergent trajectories. Then, the stability study and stabilization of a switched system with arbitrary switching is not straightforward. The following lemma is used to derive a method to determine the observer gain.

**Lemma [10]:** The autonomous switched system (10) is asymptotically stable in the origin if there exist  $n \times n$  symmetric positive matrices  $P_i$ , for  $i \in \mathfrak{I}$ , satisfying the following set of LMIs:

$$\begin{bmatrix} P_i & \mathcal{A}_i^T P_j \\ P_j \mathcal{A}_i & P_i \end{bmatrix} > 0, \quad \forall (i, j) \in \mathfrak{I} \times \mathfrak{I} \quad (11)$$

■

Substituting  $\mathcal{A}_i = \Lambda(k_t = i - 1)$  in (11), for  $i \in I = \{1, 2, \dots, N + 1\}$  leads to a Bilinear Matrix Inequality (BMI) with respect to matrix variables  $L$  and  $P_i$  for  $i \in \mathfrak{I}$ . In what follows, an algorithm is explained to solve the derived BMI.

Inspired by [11], the linearization method is employed to solve the resulting BMI. The approach is to achieve the design objective by iteratively solving a sequence of linearized problems. Substituting  $L = L_0 + \alpha_L \Delta L$  and  $P_i = P_{i0} + \alpha_P \Delta P_i$  in the obtained BMI and neglecting the second order terms results in an LMI in the variables  $\Delta L$  and  $\Delta P_i$ . As all the local methods to solve nonlinear feasibility problems, the choice of initial value  $L_0$  and  $P_{i0}$  is important for convergence to an acceptable solution. A good guess for initial  $L_0$  can be achieved by pole placement such that all the subsystems of switched system (9) are made stable. Constant coefficients  $\alpha_L$  and  $\alpha_P$  are parameters to adjust the convergence rate of iterative procedure.

#### IV. AN ILLUSTRATIVE EXAMPLE

To illustrate the validity of the proposed method, a numerical example is presented.

**Example:** Consider an open loop unstable system in the form of (1) with the following system matrices:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -5 \\ 3 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$$

The problem is solved for  $N = 2$ , so  $k_t = 0, 1, 2$  and the error system (9) is composed of three subsystems with  $\mathcal{A}_1 = \Lambda(k_t = 0)$ ,  $\mathcal{A}_2 = \Lambda(k_t = 1)$  and  $\mathcal{A}_3 = \Lambda(k_t = 2)$ . The initial observer gain matrix  $L_0$  is computed by pole placement such that all the eigenvalues of the error subsystems are within the unit circle.

$$L_0 = \begin{bmatrix} -0.3000 & 0.2000 \\ 2.2400 & -0.5600 \end{bmatrix}.$$

Although the three error subsystems are made stable by the above  $L_0$ , the switched system is not stable. Figure 1 shows the estimation error during 100 sample times for variable measurement delay. To overcome the problem,

the proposed method is utilized to obtain suitable observer gain.

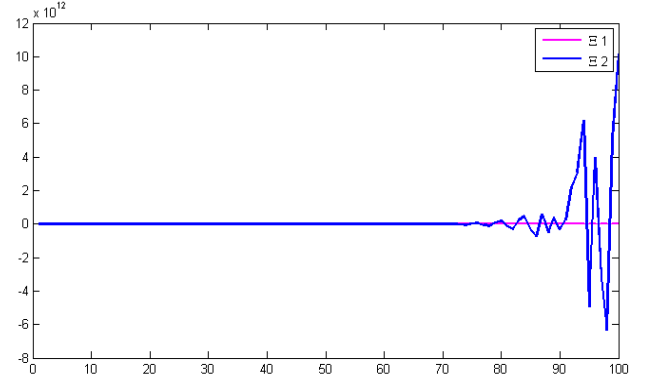


Figure 1. Observation error with gain  $L_0$

Therefore,  $\mathcal{A}_1 = \Lambda(k_t = 0)$ ,  $\mathcal{A}_2 = \Lambda(k_t = 1)$  and  $\mathcal{A}_3 = \Lambda(k_t = 2)$  is substituted in (11) and a set of 9 BMIs are obtained. To solve these BMIs simultaneously, the mentioned algorithm is carried out by setting  $\alpha_L = \alpha_P = 0.003$  and using the above  $L_0$  as initial guess of observer gain. The procedure yields to the following gain for the networked observer:

$$L = \begin{bmatrix} -0.5621 & 0.3747 \\ 1.4762 & -0.3691 \end{bmatrix}.$$

Figure 2 demonstrates the observer performance with the above achieved  $L$ .

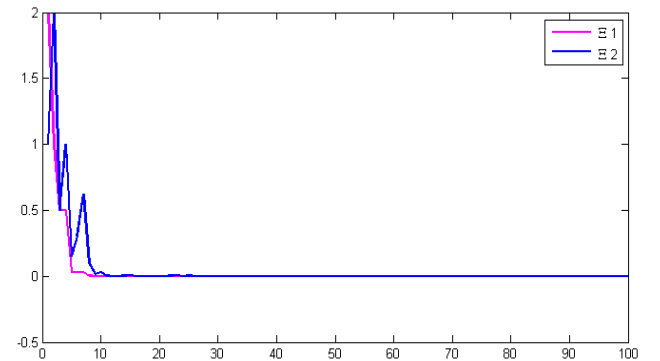


Figure 2. Observation error with gain  $L$

#### V. CONCLUSION

The Design of state observer in networked control system (NCS) was discussed. The network induced delay and data loss in the feedback channel was considered. The observer design problem was transformed to stabilization problem of a switched system. A BMI condition was extracted to determine the observer gain. Furthermore, a procedure suggested to solve the mentioned BMI. Finally, a numerical example was presented to illustrate the validity of the proposed method.

The future work, will consider the variable delays in both feedback and forward channel.

#### REFERENCES

- [1] J. P. Hespanha, P. Naghshtabrizi and Y. Xu, "A Survey of Recent Results in Networked Control Systems", *Proceedings of IEEE*, Vol. 95, No. 1, pp. 138-162, 2007.
- [2] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla and S. S. Sastry, "Foundation of Control and estimation over Lossy Networks", Vol. 95, No. 1, pp. 163- 187, 2007.
- [3] C. Lin, Z. Wang and F. Yang, "Observer-based Networked Control for Continuous-time Systems with Random Sensor Delays", *Automatica*, Vol. 45, pp. 578-584, 2009.
- [4] Dragan Dacic and Dragan Nesic, "Observer design for wired linear networked control systems using matrix inequalities", *Automatica*, Vol. 44, pp. 2840-2848, 2008.
- [5] Luca Schenato, "Optimal Estimation in Networked Control Systems Subject to Random Delay and Packet Drop", *IEEE Transactions on Automatic Control*, Vol. 53, No. 5, pp. 1311-1317, 2008.
- [6] G. P. Liu, Y. Xia, D. Rees and W. Hu, "Design and Stability Criteria of Networked Predictive Control Systems with Random Network Delay in the Feedback Channel", *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 37, No. 2, pp. 173-184, 2007.
- [7] Z. Wang, D. Ho, Y. Liu and X. Liu, "Robust H-infinity Control for a Class of Nonlinear Discrete Time-Delay Stochastic Systems with Missing Measurements", *Automatica*, Vol. 45, No. 3, pp. 684-691, 2009.
- [8] B. Shen, Z. Wang, H. Shu and G. Wei, "On Nonlinear H-infinity Filtering for Discrete-Time Stochastic Systems with Missing Measurements", *IEEE Transactions on Automatic Control*, Vol. 53, No. 9, pp. 2170-2180, 2008.
- [9] H. Dong, Z. Wang and H. Gao, "Robust H\_infinity Filtering for a Class of Nonlinear Networked Systems with Multiple Stochastic Communication Delays and Packet Dropouts", *IEEE Transactions on Signal Processing*, Vol. 58, pp. 1957-1966, 2010.
- [10] L. Fang, H. Lin, and P. J. Antsaklis, "Stabilization and Performance Analysis for a Class of Switched Systems", *43<sup>rd</sup> IEEE Conference on Decision and control*, December 14-17, 2004, Atlantis, Paradise Island, Bahamas.
- [11] A. Hassibi, J. How, S. Boyd, "A Path-following Method for Solving BMI Problems in Control," *Information Systems Laboratory Stanford University Stanford, CA* 94305-9510, USA.

# Stabilization of Networked Control Systems with Variable Transmission Delays

M. Mahmodi Kaleybar , R. Mahboobi Esfanjani

Electrical Engineering Department,  
Sahand University of Technology  
Tabriz, Iran

[M\\_Mahmodi@sut.ac.ir](mailto:M_Mahmodi@sut.ac.ir), [mahboobi@sut.ac.ir](mailto:mahboobi@sut.ac.ir)

**Abstract**— In this paper, an approach is proposed to design static state feedback controller for stabilization of networked control systems (NCSs). Both of the data packet delay and dropout which degrade the performance of the closed-loop system are considered in the NCS model. Feedback gain is determined by solving matrix inequality which is notably less complicated compared with the existing results. Furthermore, the maximum tolerable transfer interval can be derived by the suggested method. Numerical examples are introduced to demonstrate the effectiveness and advantages of this scheme.

**Keywords**- *Networked Control Systems; Variable Delay; Stabilization; State Feedback Controller.*

## I. INTRODUCTION

A Networked Control System (NCS) is a feedback control system wherein the control loop is closed through a communication network. The advantages of NCS such as low cost, simple installation and maintenance and high flexibility make it more and more popular in many applications including distributed industrial control systems, intelligent traffic systems, cluster of unmanned air vehicles and multi-agent systems. Despite of these benefits and potentials, presence of communication network in control loop make the analysis and design of the control system very complicated. Main issues in this configuration are the packet dropouts and delays in the sensor-to-controller and controller-to-actuator channels which occur when sensor, actuator and controller exchange data across the network.

Various approaches have been developed in the recent years to synthesis stabilizing controllers for linear time-invariant systems in the presence of transmission delays and packet dropouts. The problem of data packet dropout and transmission delays induced by communication channel in networked control system is studied in [1]-[6]. In [1], networked control system with random transmission delay was considered in the switched systems' framework. A predictive control scheme was proposed to compensate the effects of network delay and data dropout. In [2], the networked control system with data packet dropout was modeled as jump linear system. An approach proposed to obtain constant stabilizing state-feedback controller by solving linear matrix inequalities (LMIs). In [3], a controller design method was proposed based on a delay-dependent approach for networked control system in which both the network-induced delay

and the data packet dropout in the transmission are taken into account. The feedback gain of controller and the maximum allowable value of the network-induced delay can be derived by solving a set of LMIs. The results of [3] were improved in [4], wherein a memoryless state feedback is designed in similar manner to stabilize NCS by solving a set of LMIs. The Lyapunov-Razumikhin technique was utilized in [5] to derive a delay dependent condition for the stabilization of NCS in terms of LMIs. In [6], for a class of nonlinear discrete time-delay systems with missing measurements, a robust controller was developed.

In this note, inspired by the NCS model developed in [3] and [4], a procedure is derived to design a memoryless state feedback stabilizing controller. Utilizing an appropriate Lyapunov-Krasovskii functional [7], a delay-dependent sufficient condition is obtained in terms of LMI which is computationally more tractable compared to [3]-[5]. Maximum allowable transfer interval as a common criterion to measure the performance of the NCS is calculated for numerical examples to demonstrate the efficiency of the proposed approach.

The paper is organized as follows: In section II, networked control system model is described. Section III presents the main results of this paper wherein controller synthesis method is derived to determine controller gain and allowable upper bound of variable delay. In section IV, the suggested design method is illustrated by two examples and the results are compared with the existing design approaches. Finally, section V concludes this paper.

## II. PROBLEM SETUP

The sensor, controller and the actuator are assumed to be separated and connected through a communication network. In the considered network, the data are lumped together into one packet and transmitted at the same time (single packet transmission). In the considered NCS, the controlled system is linear and time invariant, sensor is time-driven and controller and actuator are event-driven. The sensor packet and its related control packet are stamped with the sampling time instant. The controller and actuator always use the new received data packets and discard the old ones which means when an old data packet arrives, it is dealt with as a packet loss. The input of the

plant is implemented through a zero-order-hold and is zero before the first controller packet arrives.

As depicted in Fig. 1, the transmission delays induced by the network are shown with  $\tau_{sc}$ , for sensor-to-controller delay and  $\tau_{ca}$ , for controller-to-actuator delay. In fact, if the feedback controller is static these two delays can be combined as  $\tau_{i_k} = \tau_{sc} + \tau_{ca}$ , which is a variable parameter. Regarding the above assumptions on the NCS, the following dynamical equation can describe the closed-loop system behavior:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad t \in [i_k h + \tau_{i_k}, i_{k+1} h + \tau_{i_{k+1}}) \quad (1)$$

$$\mathbf{u}(t^+) = \text{sat}(\mathbf{K}\mathbf{x}(t - \tau_{i_k})), \quad t \in \{i_k h + \tau_{i_k}, k = 1, 2, \dots\} \quad (2)$$

where  $\mathbf{x}(t) \in R^n$  and  $\mathbf{u}(t) \in R^m$  are the state vector and the control input vector respectively and  $\mathbf{A}, \mathbf{B}$  are two constant matrices with appropriate dimensions.  $h$  is the sampling period and  $i_k$  is an integer denoting the sampling instant of the state feedback for  $k = 1, 2, 3, \dots$ .  $t^+$  denotes the time interval ranging from  $i_k h + \tau_{i_k}$  to  $i_{k+1} h + \tau_{i_{k+1}}$ .  $\tau_{i_k}$  stands for the network-induced time delay from the instant  $i_k h$ .  $\mathbf{K}$  is the state feedback gain matrix will be designed.

In the time interval  $t \in [i_k h + \tau_{i_k}, i_{k+1} h + \tau_{i_{k+1}})$ , the closed-loop system model can be rewritten as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{K}\mathbf{x}(i_k h),$$

By definition of  $\tau(t) = t - i_k h$  one gets:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{K}\mathbf{x}(t - \tau(t)) \quad (3)$$

for  $t \in [i_k h + \tau_{i_k}, i_{k+1} h + \tau_{i_{k+1}})$ , in which :

$$0 \leq \tau_{i_k} \leq \tau(t) \leq (i_{k+1} - i_k)h + \tau_{i_{k+1}} \leq \bar{\tau} \quad (4)$$

where  $\bar{\tau}$  is supposed to be upper bound of tolerable delay to preserve the closed-loop stability and  $\mathbf{x}(t) = \varphi(t)$  for  $t \in [t_0 - \bar{\tau}, t_0]$  is the initial condition of the systems.

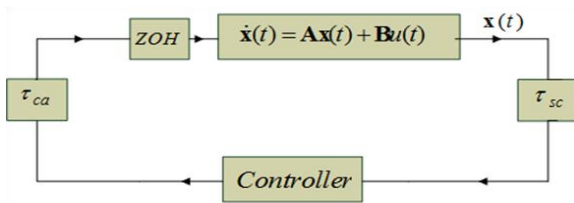


Figure 1. Schematic Diagram of the NCS

Consequently, the NCS was modeled as linear time-delay system (3) with variable random delay which represents both the data packet dropout and latency in the network. This fact allows the use of time-delay systems theory to study the NCS.

### III. CONTROLLER DESIGN

The following integral-inequality is used in the design procedure to obtain the static feedback controller:

**Lemma 1 [8]:** Let  $\mathbf{x}(t) \in R^n$  be a vector-valued function with first-order continuous derivative. For any constant matrix  $\mathbf{R} = \mathbf{R}^T > 0$  and scalar  $0 \leq \tau \leq \bar{\tau}$ , the following integral-inequality holds:

$$-\int_{t-\tau(t)}^t \bar{\tau} \dot{\mathbf{x}}^T(s) \mathbf{R} \dot{\mathbf{x}}(s) ds \leq -\xi_1^T(t) \begin{bmatrix} -\mathbf{R} & \mathbf{R} \\ * & -\mathbf{R} \end{bmatrix} \xi_1(t)$$

where  $\xi_1(t) = [\mathbf{x}(t), \mathbf{x}(t - \tau)]^T$  and the  $*$  denotes the symmetric entry in a symmetric matrix. ■

Lyapunov-Krasovskii functional candidate is considered as follows:

$$V(t) = \mathbf{x}^T(t) \mathbf{P} \mathbf{x}(t) + \int_{t-\tau}^t \mathbf{x}^T(s) \mathbf{Q} \mathbf{x}(s) ds + \bar{\tau} \int_{-\bar{\tau}}^0 \int_{t+\theta}^t \dot{\mathbf{x}}^T(s) \mathbf{R} \dot{\mathbf{x}}(s) ds d\theta \quad (5)$$

in which the matrices  $\mathbf{P} = \mathbf{P}^T > 0$ ,  $\mathbf{Q} = \mathbf{Q}^T > 0$  and  $\mathbf{R} = \mathbf{R}^T > 0$  will be determined. The sufficient condition to determine the mentioned matrices and feedback gain is given in the following theorem.

**Theorem 1.** For given scalars  $h$  and  $\bar{\tau}$  and  $\dot{\tau}(t) \leq d$ , if there exist matrices  $\mathbf{P} = \mathbf{P}^T > 0$ ,  $\mathbf{Q} = \mathbf{Q}^T > 0$ ,  $\mathbf{R} = \mathbf{R}^T > 0$  and  $\mathbf{S}_i$  ( $i = 1, 2, 3$ ) of appropriate dimensions such that (6) holds, then the system (1) with the state feedback controller (2) and variable delay satisfying (4) is stable:

$$\Phi < 0 \quad (6)$$

where  $\Phi = \Omega_1 + \Omega_2 + \Omega_2^T$ .

$$\Omega_1 = \begin{bmatrix} \mathbf{Q} - \mathbf{R} & \mathbf{R} & \mathbf{P} \\ * & -(1-d)\mathbf{Q} - \mathbf{R} & 0 \\ * & * & \bar{\tau}^2 \mathbf{R} \end{bmatrix}$$

$$\Omega_2 = [\mathbf{S} \mathbf{A} \quad -\mathbf{S} \mathbf{B} \mathbf{K} \quad \mathbf{S}]$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \mathbf{S}_3 \end{bmatrix}$$

**Proof:** Calculating the time derivative of  $V(t)$  with respect to  $t$  along the trajectories of the system (3) for  $t \in [i_k h + \tau_{i_k}, i_{k+1} h + \tau_{i_{k+1}})$  yields:

$$\begin{aligned} \dot{V}(t) &= 2\mathbf{x}^T(t) \mathbf{P} \dot{\mathbf{x}}(t) + \mathbf{x}^T(t) \mathbf{Q} \mathbf{x}(t) \\ &\quad - (1 - \dot{\tau}(t)) \mathbf{x}^T(t - \tau) \mathbf{Q} \mathbf{x}(t - \tau) \\ &\quad + \bar{\tau}^2 \dot{\mathbf{x}}^T(t) \mathbf{R} \dot{\mathbf{x}}(t) - \bar{\tau} \int_{t-\bar{\tau}}^t \dot{\mathbf{x}}^T(s) \mathbf{R} \dot{\mathbf{x}}(s) ds \end{aligned} \quad (7)$$



Note that, the following equation holds for any matrix  $\mathbf{S}$  and vector  $\xi(t) = [\mathbf{x}^T(t) \quad \mathbf{x}^T(t-\tau) \quad \dot{\mathbf{x}}^T(t)]^T$ :

$$2\xi^T(t)\mathbf{S}[\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) - \mathbf{B}\mathbf{K}\mathbf{x}(t-\tau)] = 0$$

Regarding the above relation and Lemma 1, the following inequality presents an upper bound for the  $\dot{V}(t)$  in (7):

$$\begin{aligned} \dot{V}(t) \leq & 2\mathbf{x}^T(t)\mathbf{P}\dot{\mathbf{x}}(t) + \mathbf{x}^T(t)\mathbf{Q}\mathbf{x}(t) \\ & - (1-d)\mathbf{x}^T(t-\tau)\mathbf{Q}\mathbf{x}(t-\tau) + \bar{\tau}^2\dot{\mathbf{x}}^T(t)\mathbf{R}\dot{\mathbf{x}}(t) \\ & - [\mathbf{x}(t) \quad \mathbf{x}(t-\tau)] \begin{bmatrix} -\mathbf{R} & \mathbf{R} \\ \mathbf{R} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}(t-\tau) \end{bmatrix} \\ & + 2\xi^T(t)\mathbf{S}[\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) - \mathbf{B}\mathbf{K}\mathbf{x}(t-\tau)] \end{aligned} \quad (8)$$

By definition of  $\Omega_1$  and  $\Omega_2$ , the following is obtained:

$$\dot{V}(t) \leq \xi^T(t)(\Omega_1 + \Omega_2 + \Omega_2^T)\xi(t). \quad (10)$$

If (6) holds, it is clear that  $\dot{V}(t) \leq \xi^T(t)\Phi\xi(t) < 0$ , where  $\Phi = \Omega_1 + \Omega_2 + \Omega_2^T$ . Now, suppose that  $\lambda = \lambda_{\min}(-\Phi) > 0$ , the following inequality is true:

$$\begin{aligned} \dot{V}(t) & < -\lambda\|\mathbf{x}(t)\|^2 - \lambda\|\mathbf{x}(t-\tau)\|^2 - \lambda\|\dot{\mathbf{x}}(t)\|^2 \\ & < -\lambda\|\mathbf{x}(t)\|^2 \end{aligned}$$

Also, it is obvious that there is positive scalar  $\varepsilon$  such that  $V(t) > \varepsilon\|\mathbf{x}(t)\|^2$ . So, by the Lyapunov-Krasovskii Theorem [7], the closed-loop system is stable. ■

**Remark 1:** Theorem 1 gives the sufficient condition for the stability of the system (1) with the networked memoryless state feedback controller (2). For a given controller gain  $\mathbf{K}$ , Theorem 1 can be used to compute the maximum value of  $\bar{\tau}$ .

The inequality (6) is a nonlinear matrix inequality, so it cannot be solved by efficient LMI methods. In the following theorem, based on changing variable method and parameter tuning, the controller design condition is modified to obtain an LMI condition.

**Theorem 2:** For given scalars  $h$ ,  $\bar{\tau}$  and  $p_i$  ( $i=1,2$ ), if there exist matrices  $\tilde{\mathbf{P}} = \tilde{\mathbf{P}}^T > 0$ ,  $\tilde{\mathbf{Q}} = \tilde{\mathbf{Q}}^T > 0$ ,  $\tilde{\mathbf{R}} = \tilde{\mathbf{R}}^T > 0$ ,  $\tilde{\mathbf{S}}_i$  ( $i=1,2,3$ ) and nonsingular matrices  $\mathbf{X}$  and  $\mathbf{Y}$  of appropriate dimensions such that the LMI (11) holds, then the system (1) with the networked static state feedback controller (2) is stable.

$$\tilde{\Phi} < 0 \quad (11)$$

where  $\tilde{\Phi} = \tilde{\Omega}_1 + \tilde{\Omega}_2 + \tilde{\Omega}_2^T$ .

$$\tilde{\Omega}_1 = \begin{bmatrix} \tilde{\mathbf{Q}} - \tilde{\mathbf{R}} & \tilde{\mathbf{R}} & \tilde{\mathbf{P}} \\ * & -(1-d)\tilde{\mathbf{Q}} - \tilde{\mathbf{R}} & 0 \\ * & * & \bar{\tau}^2\tilde{\mathbf{R}} \end{bmatrix}$$

$$\begin{aligned} \tilde{\Omega}_2 &= \begin{bmatrix} \mathbf{A}\mathbf{X}^T & -\mathbf{B}\mathbf{Y} & \mathbf{X}^T \\ p_2\mathbf{A}\mathbf{X}^T & -p_2\mathbf{B}\mathbf{Y} & p_2\mathbf{X}^T \\ p_3\mathbf{A}\mathbf{X}^T & -p_3\mathbf{B}\mathbf{Y} & p_3\mathbf{X}^T \end{bmatrix} \\ \tilde{\mathbf{S}} &= \begin{bmatrix} \tilde{\mathbf{S}}_1 \\ \tilde{\mathbf{S}}_2 \\ \tilde{\mathbf{S}}_3 \end{bmatrix}. \end{aligned}$$

Furthermore, the state feedback control gain is obtained by  $\mathbf{K} = \mathbf{Y}\mathbf{X}^{-T}$ .

**Proof :** In the matrix  $\mathbf{S}$  defined in (6), replace  $\mathbf{S}_1 = \mathbf{S}_0$ ,  $\mathbf{S}_2 = p_2\mathbf{S}_0$  and  $\mathbf{S}_3 = p_3\mathbf{S}_0$ . Feasibility of inequality (6) implies that  $\mathbf{S}_0$  is nonsingular. Let  $\mathbf{X} = \mathbf{S}_0^{-1}$ , then pre and post multiply simultaneously the two sides of (6) with  $\text{diag}\{\mathbf{X}, \mathbf{X}, \mathbf{X}, \mathbf{X}\}$  and its transpose, respectively. Set  $\tilde{\mathbf{P}} = \mathbf{X}\mathbf{P}\mathbf{X}^T$ ,  $\tilde{\mathbf{Q}} = \mathbf{X}\mathbf{Q}\mathbf{X}^T$ ,  $\tilde{\mathbf{R}} = \mathbf{X}\mathbf{R}\mathbf{X}^T$ ,  $\mathbf{Y} = \mathbf{K}\mathbf{X}^T$  and  $\tilde{\mathbf{S}}_i = \mathbf{X}\mathbf{S}_i\mathbf{X}^T$ , ( $i=1,2,3$ ). Therefore, inequality (6) leads to inequality (11). ■

**Remark 2:** Theorem 2 gives a condition to obtain the state feedback gain  $\mathbf{K}$  in terms of LMI for the system (3). However, using the tuning parameters in Theorem 2 leads to some conservativeness in the resulting sufficient condition compared to Theorem 1. Therefore, the result of Theorem 1 which is an LMI with specified  $\mathbf{K}$ , is used to compute the maximum allowable value of  $\bar{\tau}$  to achieve less conservative bound.

**Remark 3:** The  $\tilde{\Phi}$  in inequality (11) is a  $3n \times 3n$  matrix, where  $n$  is the number of system states. While the corresponding condition in [4] involves a matrix of dimension  $7n \times 7n$ . So, the proposed method is computationally more tractable than the one in [4].

#### IV. NUMERICAL EXAMPLES

To illustrate the effectiveness of the proposed method, we present two numerical examples.

**Example 1:** Consider the simplified model of the inverted system process [4]:

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathbf{u}(t) \quad (12)$$

Utilizing YALMIP Toolbox [9], the maximum admissible value of  $\bar{\tau}$  is obtained 0.994 with the state feedback gain  $\mathbf{K} = [-1.0056 \quad -1.0056]$  which is close to the results reported in [4]. However, Computational burden of the proposed method is about one third of the method in [4]. Fig.2 demonstrates the state response of the controlled system (11) for a typical scenario of variable delay.

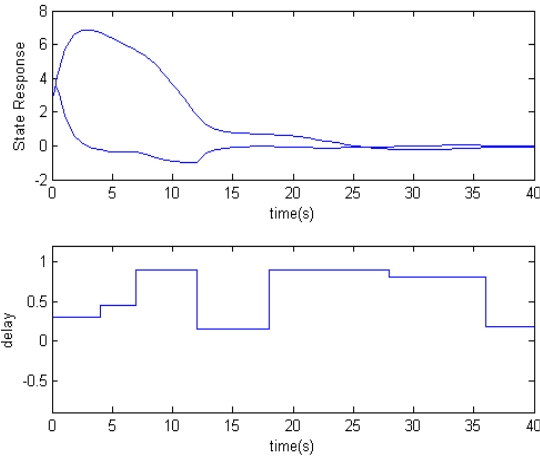


Figure 2. State Response of Networked Controlled System (12) with the Variable Delay Scenario

**Example 2:** Consider the system

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -0.1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} \mathbf{u}(t) \quad (13)$$

The networked controller gain is determined by the proposed method and the corresponding maximum allowable transfer intervals (MATI) which is also called MADB [9], is computed. Table 1 lists the MATIs corresponding to the proposed design methods for different values of  $d$  ( the upper bound of delay variation rate ).

TABLE I. THE MATI FOR DIFFERENT VALUES OF  $d$

$d$	0	0.1	0.2	0.35	0.6792
MATI	1.075	1.0571	1.0381	1.008	1.0081

For easy comparison, Table 2 shows the MATI from the references [3], [4] and [11-13] for the controlled system (11). It's clear that if the network delay variation is less than 0.6792, the proposed method results in the MATI which is comparable to best result in the literature, while the computation complexity is very low.

TABLE II. MATI IN DIFFERENT REFERENCES

Ref.	[13]	[10]	[3]	[11]	[4]
MATI	0.00045	0.0583	0.8695	0.9410	1.0081

## I. CONCLUSION

This paper proposed an approach to design stabilizing state feedback controller for the linear time invariant system which is controlled via communication network. Data packet loss and latency arising in NCS were modeled as time varying delay and the upper bound of the admissible delay to retain the closed-loop stability

was computed. In the suggested scheme stabilizing state feedback controller is constructed via the feasible solution of a matrix inequality which is considerably less complicated than existing approaches. The applicability of the method was demonstrated by numerical examples.

## REFERENCES

- [1] G. P. Liu, Y. Xia, J. Chen, D. Rees and W. Hu, "Networked predictive control of systems with random network delays in both forward and feedback channels," IEEE Transactions on Industrial Electronics, vol.54, pp.1282-1297, 2007.
- [2] M. Yu, L. Wang, T. Chu and G. Xie, "Modeling and control of networked systems via jump systems approach," IET Control Theory and Applications, vol.2, pp.535-541, 2008.
- [3] D. Yue, Q.L. Han, and C. Peng, "State feedback controller design of networked control systems," IEEE Transactions on Circuits and Systems II : Express Briefs, vol.51, pp.640-644, 2004.
- [4] Bin Tang, Guo-Ping Liu and Wei-Hua Gui, "Improvement of state feedback controller design for networked control systems," IEEE Transactions on Circuits and Systems, vol.55, pp.464-468, 2008.
- [5] M. Yu, L. Wang, T. Chu and F. Hao, "Stabilization of networked control systems with data packet dropout and transmission delays: continuous-time case," European Journal of Control, vol.11, pp.40-49, 2005.
- [6] Z. Wang, D. Ho, Y. Liu and X. Liu, "Robust H-infinity control for a class of nonlinear discrete time-delay stochastic systems with missing measurements," Automatica, Vol. 45, pp. 684-691, No. 3, Mar 2009.
- [7] J.K. Hale and S.M. Verduyn Lunel, "Introduction to Functional Differential Equations," Volume 99 of Applied Mathematical Sciences, Springer Verlag, 1993.
- [8] C. Peng, Y. Tian and M. O. Tade, "State feedback controller design of networked control systems with interval time-varying delay and nonlinearity", International Journal of Robust and Nonlinear control, vol.10, pp.1-16, 2007.
- [9] J. Löfberg, YALMIP: A toolbox for modeling and optimization in MATLAB Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.
- [10] Y. He, Q.G. Wang, L. Xie, and C. Lin, "Further improvement of free weighting matrices technique for system with time-varying delay," IEEE Trans. on Automatic Control, vol.52, pp. 293-299, 2007.
- [11] G.C. Walsh, H. Ye and L.G. Bushnell, "Stability analysis of networked control systems," IEEE Trans. on Control Systems Technology, vol.10, pp.438-446, 2002.
- [12] H.S. Park, Y.H. Kim, D.S. Kim and W.H. Kwon, "A scheduling method for network based control systems," IEEE Trans. on Control Systems Technology, vol.10, pp.318-330, 2002.
- [13] D.S. Kim, T.S. Lee, W.H. Kwon and H.S. Park, "Maximum allowable delay bounds of networked control systems," Control Engineering Practice, vol.11, pp.1301-1313, 2003.

# Zero Overshoot and Fast Transient Response Using a Fuzzy Logic Controller

Bakhtiar I. Saeed & Bruce Mehrdadi  
School of Computing and Engineering,  
University of Huddersfield  
Queensgate, Huddersfield HD1 3DH, UK  
[b.saeed@hud.ac.uk](mailto:b.saeed@hud.ac.uk), [b.mehrdadi@hud.ac.uk](mailto:b.mehrdadi@hud.ac.uk)

**Abstract**— In some industrial process control systems it is desirable not to allow an overshoot beyond the setpoint or a threshold, this could be a safety constraint or the requirement of the system. This paper outlines our work in designing a fuzzy PID controller to achieve a step-response with zero overshoot while improving the output transient response. Our designed fuzzy PID controller is applied to stable, marginally stable and unstable systems and their step responses are compared with a tuned conventional PID controller. A comparative case study shows that the proposed fuzzy controller is highly effective and outperforms the PID controller in achieving a zero overshoot response and enhancing the output transient response.

**Keywords** - fuzzy PD+I; PID controller; zero overshoot; scaling gain; tuning.

## I. INTRODUCTION

An integral part in controller design and analysis is to achieve a satisfactory response in transient time and steady state. The characteristics of these states can be represented in parameters such as: overshoot, rise time, settling time and steady state error. In a stable system, the transient response exists for a short period of time. However, this might cause problems in some applications. For example, in some chemical processes, it is desirable to have zero overshoot or an overshoot that does not exceed a specific threshold.

The conventional Proportional-Integral-Derivative (PID) controllers, which are the most popular feedback methods for their robustness and simplicity [1, 2] have some limitations, particularly when they are applied to obtain zero overshoot. These controllers can be tuned in several ways [3] to achieve zero overshoot if possible, but most of the time this is at the expense of rise time, and vice versa.

Some methods have been reported by researchers to find the values of PID gains to achieve zero overshoot [4-7]. By relating the step response overshoot to the positions of zeros and poles of a transfer function, a method has been derived to find the parameters of PID controller [6]. This method has been used to avoid overshoot in second order and lower order systems. A cascade sliding mode-PID controller has been proposed in literature [7].

On the other hand fuzzy logic controllers have been applied successfully in industrial processes and in some cases outperform PID controllers [8], in particular when

the controlled system is complex or non-linear, as this is the case in many process control systems [9].

Nonetheless, designing fuzzy controllers is challenging. There is no systematic process for the design of fuzzy logic controllers that will produce a high-performance controller for a wide range of applications [10, 11]. For example, it is difficult to find the relation between selecting membership function type or rule base, and the controller performance such as better rise time or less overshoot. In addition, unlike conventional controllers, fuzzy controllers have several parameters that can be adjusted, such as membership function shape, rules and scaling gains. Furthermore, there is no general rule of tuning these parameters. However, some techniques applied in tuning conventional controllers can still be utilised to some extent [10].

In this study, a fuzzy PID controller is adopted and is applied to different second order systems. Initially the controller gains are fixed and then manually tuned to achieve zero overshoot with a short rise time and settling time. A case study has been used to compare the performance of the fuzzy PID and conventional PID controllers for a second order system. The results show that fuzzy controllers outperform conventional controllers in achieving zero overshoot and fast transient response.

The remainder of this paper is organised as follows: Section 2 presents an overview of the fuzzy controller. Design and synthesis of the fuzzy PD+I controller is illustrated in section 3 using MATLAB and Simulink. Simulation results are shown in section 4. Finally, some conclusions are drawn in section 5.

## II. STRUCTURE OF THE FUZZY CONTROLLER

The most widely used fuzzy controllers are fuzzy proportional-derivative controllers (FPD) and they act on two inputs: error and change in error (derivative of error) signals; therefore, designing rule base for these controllers is well understood and a straightforward procedure. This configuration exhibits a good performance at the transient response of the system, while encountering problems at the steady state when the error is close to zero [3, 12, 13].

To enhance the steady state performance of a system, integral action is required [3]. Thus the controller becomes fuzzy proportional-integral controller (FPI). Although these controllers have good performance, at the steady state they suffer from a slow response [12, 13].

Improving both the transient state and the steady state requires a controller that includes both derivative and integral actions. A fuzzy controller with this capability is known as fuzzy proportional-derivative-integral controller (FPID).

In literature, various structures have been proposed to design FPID controllers [3, 13-15].

Fig.1 shows a simple design proposed in [3] and was adopted as the fuzzy PID controller in this study.

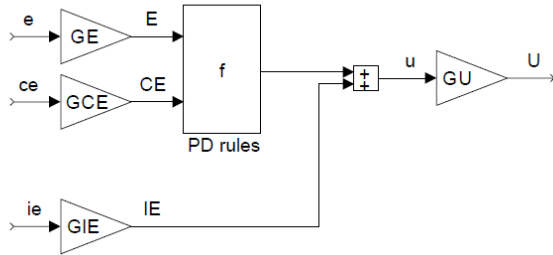


Figure 1: Fuzzy PD+I controller (FPD+I) [3]

The controller consists of a normal FPD with added integral action; therefore it is known as FPD+I. As the controller has three inputs: error, derivative of error and integral of error, it can provide all the benefits of conventional PID controllers, but still has some disadvantages such as derivative kick and integrator windup. Additionally, there is only one rule base with two inputs; therefore, designing the rule base is less complex than the structure proposed in [15] which has three input rule base. Furthermore, some techniques applied in tuning conventional PID controllers can still be utilised to some extent [3, 10].

### III. SIMULATION OF CONTROLLERS

#### A. Fuzzy PD+I

Matlab (v7.9) and Simulink were used to build and simulate the model. Fig. 2 shows the FPD+I controller in a closed-loop feedback system.

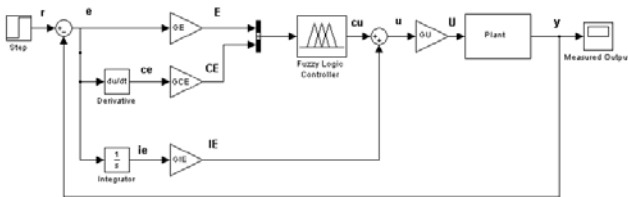


Figure 2: Simulink model of Fuzzy PD+I controller in a closed-loop control structure.

The plant block represents the desired transfer function to be controlled.

The fuzzy PD+I controller was formed by adding the integrator to the output of the fuzzy PD controller. The controller has three input signals: error (e), change in error (ce) and integral of error (ie). The error signal is obtained from the difference between the setpoint (r) and the measured plant output (y), the change in error signal and

the integral of error signals are produced by passing the signals through derivative and integral blocks respectively. The inputs have scaling gains: gain of error (GE), gain of change in error (GCE) and gain of integral error (GIE). These gains along with the output gain (GU) can be tuned to achieve better performance [3, 10, 15-17]. Adjusting these gains is used frequently for tuning fuzzy controllers and it has been regarded as an effective approach [12, 15, 16, 18-21]. First of all, they have a global effect on the performance of the controller; their effects can be easily observed [18]. Secondly, there are few parameters to tune, thus the tuning process is computationally efficient in contrast with other methods, where there are several parameters to tune. Finally, they can be considered as conventional controller gain parameters; therefore, they are convenient to tune and some ideas from conventional controller tuning can be borrowed [12, 16, 18-21].

The controller output (u) is formed by adding the integral of error (ie) to the output of the fuzzy PD controller (cu).

To represent the values of the e, ce and cu, five symmetric triangle shape membership functions (except trapezoid for the two at the extreme ends for e and ce) with 50% of overlap were chosen [3, 10]. Although the choice of membership function shape and width is subjective, triangular shapes were chosen, because they are more popular and convenient [10, 11]. The interval of  $[-1, 1]$  was used for the universes of discourse of the input variables, while  $[-2, 2]$  was used for the output variable. The linguistic descriptions of the input and output membership functions are negative large (NL), negative small (NS), zero (ZE), positive small (PS) and positive large (PL). These are shown in Fig. 3 and Fig. 4 respectively.

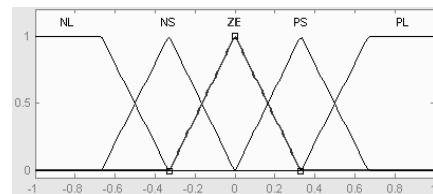


Figure 3: Error and change of error membership functions

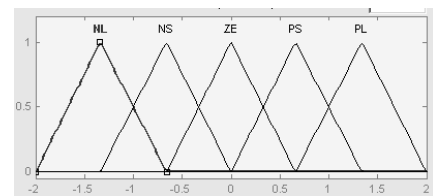


Figure 4: Output membership functions

The minimum operator was selected as an implication method, and the most popular and standard method of defuzzification process known as the centre of gravity (CoG) was selected.

The fuzzy rule-base is a mapping between the inputs,  $e$  and  $ce$  and the output,  $cu$ . A sample of the rule has the following form:

*If error is PL and change in error is PL, then control signal is PL*

The rule implies that if the error is positive large (measured output far away from the set point) and the change of error is positive large, then the control signal should be positive large to return back the output near the setpoint.

As there are 5 linguistic variables for each input, 25 rules were created using Fuzzy Logic Toolbox, Table I shows the rules.

TABLE I. RULE-BASE FOR THE FUZZY CONTROLLER

$\begin{matrix} ce \\ e \end{matrix}$	NL	NS	ZE	PS	PL
NL	NL	NL	NS	NS	ZE
NS	NL	NS	NS	ZE	PS
ZE	NS	NS	ZE	PS	PS
PS	NS	ZE	PS	PS	PL
PL	ZE	PS	PS	PL	PL

Three standard closed-loop performance criteria were chosen as design specifications to measure the performance of the controller [2]: maximum percentage overshoot ( $Mp$ ), rise time ( $tr$ ) and settling time ( $ts$ )

Finally, a script code was developed to simulate the model, calculate the  $Mp$ , the  $tr$  and the  $ts$  and generate the required plots.

#### B. PID controller

In order to compare the performance of the FPD+I controller with a conventional PID controller, a simulation model of a PID controller with auto tuning capability was created, this is shown in Fig. 5.

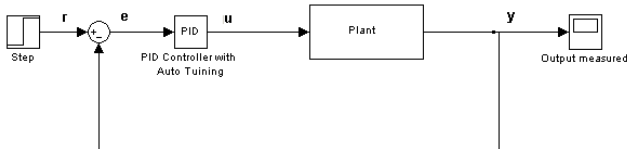


Figure 5: Simulink model of conventional PID controller in a closed-loop control structure.

The model contains a plant block that represents the desired transfer function to be controlled and a PID controller block with auto tuning capability. Also for this design, the same performance measures were chosen and a script code was developed to simulate the model.

## IV. RESULTS AND DISCUSSIONS

### A. Tests

In order to evaluate the design, three different systems: stable, marginally stable and unstable were simulated using the FPD+I and the conventional PID models. The transfer functions of these systems are provided in Table II.

TABLE II. DIFFERENT SYSTEM TRANSFER FUNCTIONS

System	Transfer Function	Stability
1	$\frac{1}{s^2 + 0.5s + 1}$	Stable
2	$\frac{1}{s^2 + 1}$	Marginally stable
3	$\frac{1}{s^2 - 15s + 1}$	Unstable

Initially, the gains of the FPD+I controller  $GE$ ,  $GCE$ ,  $GIE$  and  $GU$  are set to 1, and then tuned to achieve zero overshoot with a fast rise time and short settling time, the tuned values are shown in Table III.

TABLE III. FPD+I CONTROLLER TUNED GAIN VALUES

System	GE	GCE	GIE	GU
1	1	0.6	0.25	5
2	1	0.4	0.15	12
3	0.8	0.025	0.01	2800

For the conventional PID controller, the Matlab PID auto tuner was used to obtain the values of PID gains ( $P$ ,  $I$  and  $D$ ), these values are shown in Table IV.

TABLE IV. CONVENTIONAL PID CONTROLLER GAIN VALUES

System	P	I	D
1	11.74	0.85	8.85
2	12.04	1.09	12.088
3	3273.19	6345.72	325.43

The step responses of the three systems for the FPD+I and the conventional PID are combined together and shown in the Fig. 5 - Fig. 7.

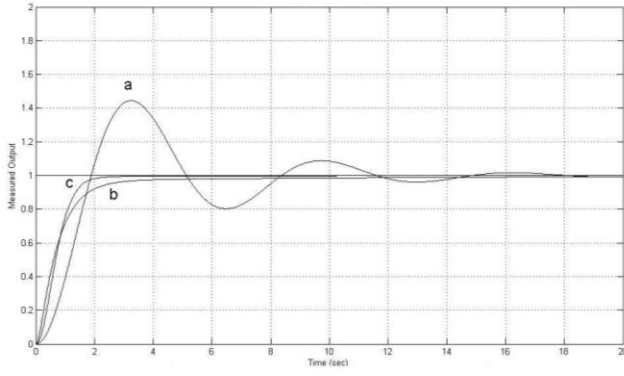


Figure 6: Simulation results for the first system: (a) open-loop. (b) Conventional PID (c) Fuzzy PD+I.

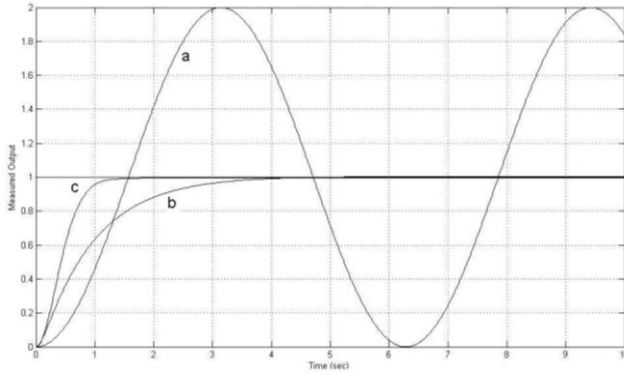


Figure 7: Simulation results for the second system: (a) open-loop. (b) Conventional PID (c) Fuzzy PD+I.

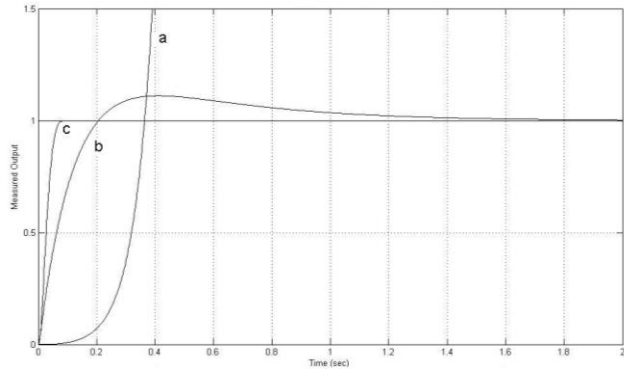


Figure 8: Simulation results for the third system: (a) open-loop. (b) Conventional PID (c) Fuzzy PD+I.

The performance measures of each controller are shown in Table V.

The performance measures of each controller and systems.

System	Performance measure	Open-loop	Conventional PID	Fuzzy PD+I
1	$M_p$	% 44.43	% 0	% 0.0
	$t_r$	1.25	1.63	1.10
	$t_s$	14.11	5.92	1.97
2	$M_p$	% 100.0	% 0.43	% 0.0
	$t_r$	1.01	1.97	0.67
	$t_s$	Not known	3.42	1.16
3	$M_p$	% 1.28 e+19	% 11.12	% 0.0
	$t_r$	0.13	0.14	0.041
	$t_s$	Not known	1.22	0.06

The results obtained from the fuzzy PD+I clearly indicate substantial improvements in transient response of systems have been achieved.

### B. Case study

The transfer function of a chemical process shown in (1) has been used by other researchers [6] to achieve zero overshoot in the closed-loop response. Accordingly, the parameters of the PID controller were calculated ( $P=7.2$ ,  $I=0.972$  and  $D=6.99$ ).

$$G(s) = \frac{1}{s(1+s)(1+10s)} \quad (1)$$

Three tests were conducted on the above system: the FPD+I controller with tuned gains, a conventional PID controller using the parameters proposed by the method in [6] and the conventional PID controller using Matlab auto tuner. The parameters of the three controllers were as follows: FPD+I ( $GE = 1$ ,  $GCE = 0.7$ ,  $GIE = 0.04$ ,  $GU = 20$ ), PID controller using the method in [6] ( $P = 7.2$ ,  $I = 0.72$ ,  $D = 6.99$ ) (The original values were for  $K_C$ ,  $T_i$  and  $T_d$ , they were converted to the values of  $P$ ,  $I$  and  $D$  to be used within the setting of Matlab PID controller) and for the conventional PID controller using Matlab auto tuner ( $P = 1.74$ ,  $I = 0.20$ ,  $D = -4.58$ ).

The closed-loop step responses of the three controllers along with the open-loop step response are shown in Fig. 9.

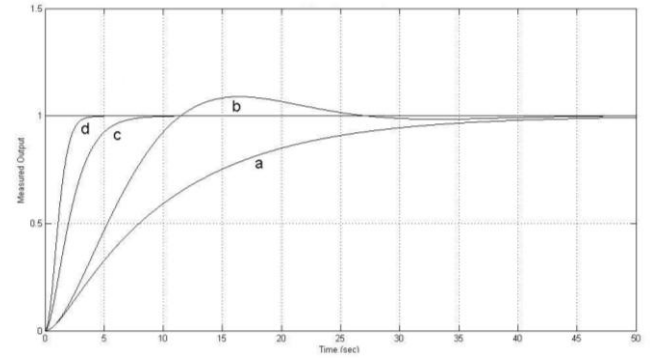


Figure 9: Simulation results: (a) open-loop. (b) Conventional PID (parameters found using Matlab auto tuner). (c) Conventional PID (parameters found using the method in [6]). (d) FPD+I.

The performance measures of the controller are shown in Table VI.

The performance measures of each controller.

PM	Open-loop	Conventional PID (parameters found using the method in [6])	Fuzzy PD+I	Conventional PID (parameters found using Matlab auto tuner)
$M_p$	% 0.0	% 0.0	% 0.0	% 8.92
$t_r$	22.14	3.96	1.77	7.8546
$t_s$	40.17	6.90	3.05	24.53



### C. Discussions

It is evident from the results, that the FPD+I controller has achieved zero overshoot with faster rise time and shorter settling time compared to the conventional PID controller. Although the conventional PID controller has achieved zero overshoot for the first and second systems, it has been at the expense of the rise time and the settling time. Additionally, in the third system and in the case study the performance of the conventional PID controller was degraded as the response resulted with some overshoot.

### V. CONCLUSIONS

A fuzzy PID controller was applied to a stable, marginally stable and unstable second order system. The results showed that fuzzy PID controller outperformed the conventional PID controllers in achieving zero overshoot and produced a faster transient response. This is an ongoing research and the next phase of the work will encompass the ability to fine tune the fuzzy gains automatically. Inclusion of predictive and intelligent agents in the fuzzy algorithms will reduce the tuning time. Although the overall controlled systems presented in this paper proved to be stable a generic stability criteria will also be incorporated in the structure of the fuzzy system.

### REFERENCES

- [1] K. J. Åström and B. Wittenmark, *Computer-controlled systems*. Upper Saddle River, N.J: Prentice Hall, 1997.
- [2] R. C. Dorf and R. H. Bishop, *Modern control systems*. Upper Saddle River: Prentice Hall, 2001.
- [3] J. Jantzen, *Foundations of fuzzy control*. Chichester: Wiley, 2007.
- [4] Y. Aiping, "A Fast generalized predictive control algorithm with non-overshoot", *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*, 2010, pp. 3572-3575.
- [5] Y.-S. Lu and C.-M. Cheng, "Design of a Non-Overshooting PID Controller with an Integral Sliding Perturbation Observer for Motor Positioning Systems", *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, vol. 48, pp. 103-110, 2005.
- [6] A. Rachid and C. Scali, "Control of overshoot in the step response of chemical processes", *Computers & Chemical Engineering*, vol. 23, pp. S1003-S1006, 1999.
- [7] Q. P. H. Thanh H. Tran, and Hung T. Nguyen, "Robust Non-Overshoot Time Responses Using Cascade Sliding Mode-PID Control", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 11, 2007.
- [8] S. R. Vaishnav and Z. J. Khan, "Design and Performance of PID and Fuzzy Logic Controller with Smaller Rule Set for Higher Order System", *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, 2007, pp. 24-26.
- [9] R. Babuska and E. Mamdani, "Fuzzy Control", [http://www.scholarpedia.org/article/Fuzzy\\_control](http://www.scholarpedia.org/article/Fuzzy_control), 2008.
- [10] K. M. Passino and S. Yurkovich, *Fuzzy control*. Menlo Park, Calif: Addison-Wesley, 1998.
- [11] I. H. Altas and A. M. Sharaf, "A generalized direct approach for designing fuzzy logic controllers in Matlab/Simulink GUI environment", *International Journal of Information Technology and Intelligent Computing*, 2007.
- [12] H. Y. Chung, B. C. Chen, and J. J. Lin, "A PI-type fuzzy controller with self-tuning scaling factors", *Fuzzy Sets and Systems*, vol. 93, pp. 23-28, 1998.
- [13] B. Subudhi, B. A. Reddy, and S. Monangi, "Parallel structure of fuzzy PID controller under different paradigms", in *Industrial Electronics, Control & Robotics (IECR), 2010 International Conference on*, 2010, pp. 114-121.
- [14] O. Karasakal, M. Guzelkaya, I. Eksin, and E. Yesil, "Online rule weighting of fuzzy PID controllers", in *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pp. 1741-1747.
- [15] C. X. Yung C. Shin, *Intelligent Systems: Modeling, Optimization, and Control*: CRC Press. Taylor & Francis Group, 2009.
- [16] X. Jie, L. Liyun, Derong, C. Yanbo, and W. Shiyu, "Fuzzy gain based adaptive fuzzy logic controller for BLDCM drive", *Control Conference, 2008. CCC 2008. 27th Chinese*, 2008, pp. 159-163.
- [17] L.-X. Wang, *Adaptive fuzzy systems and control*. Englewood Cliffs, N.J: PTR Prentice Hall, 1994.
- [18] S. Chopra, R. Mitra, and V. Kumar, "Auto Tuning of Fuzzy PI Type Controller Using Fuzzy Logic", *INTERNATIONAL JOURNAL OF COMPUTATIONAL COGNITION (HTTP://WWW.IJCC.US)*, vol. 6, 2008.
- [19] M. N. S. Melba Mary.P, Albert Singh. N., "Design of Intelligent Self-Tuning Temperature Controller for Water Bath Process", *INTERNATIONAL JOURNAL OF IMAGING SCIENCE AND ENGINEERING (IJISE)*, vol. VOL.1, 2007.
- [20] M. Murad, K. C. Cheok, and M. Das, "Methodology to simplify the tuning process of self-organizing fuzzy logic controllers", *Intelligent Engineering Systems, 2009. INES 2009. International Conference on*, 2009, pp. 57-60.
- [21] J. Victor and A. Dourado, "Adaptive scaling factors algorithm for the fuzzy logic controller", *Fuzzy Systems, 1997., Proceedings of the Sixth IEEE International Conference on*, 1997, pp. 1021-1026 vol.2.

# Trajectory Generation for Autonomous Soaring UAS

J. H. A. Clarke and W-H. Chen

Department of Aeronautical and Automotive Engineering,  
Loughborough University  
Loughborough, UK  
J.H.A.Clarke@lboro.ac.uk

**Abstract**— As unmanned aerial vehicles are expected to do more and more advanced tasks, improved range and persistence is required. This paper presents a method of using shallow layer cumulus convection to extend the range and duration of small UAVs. A simulation model of an X-Models XCalibur electric motor-glider is used in combination with a refined parametric thermal model to simulate soaring flight. The parametric thermal model builds on previous successful models with refinements to more accurately describe the weather in northern Europe. The implementation of the variation of the MacCready setting is discussed. Methods for generating efficient trajectories are evaluated and recommendations are made regarding implementation.

*UAV, UAS, Soaring, Trajectory, thermal, Optimal, Heuristic*

## I. INTRODUCTION

Over the last two decades the use of unmanned aerial vehicles (UAV) has exploded. As the use of UAVs has increased the demands placed upon the platforms have also increased. Simultaneously people desire greater access to flying assets lower down the chains of command; whether that is for military purposes or civilian survey work. This requirement necessitates the use of smaller aircraft without loss of performance. Typically the limiting factor for these small UAVs is short flight duration, limited range and payload. Many activities such as forest fire monitoring, border patrol, atmospheric research, communication relays and other surveillance tasks require greater persistence from the airframe used. Although advancements in engine and battery technology, along with miniaturisation of much of the on-board systems, continue to provide performance and capability improvements there is still a need for the introduction of novel methods to improve the range and persistence of the aircraft. One such novel solution is the extraction of energy from naturally occurring phenomena such as atmospheric convection.

Techniques to extract energy from shallow layer cumulus convection have been employed by full-size glider pilots to increase their range and duration for nearly 100 years. These soaring techniques have historically been ignored by the surveillance community because the differences in aircraft wing loading, operating speeds and efficiency rendered them pointless.

However with the latest generation of UAV this is no longer the case.

Although soaring techniques have been investigated from before the 1930s there application to UAV is a relatively new field. J. Wharington[1] was the first to propose that autonomous soaring could be a viable method for extending UAV performance (range, endurance and usable payload capacity) in 1998. Since Wharington first proposed that static soaring was a viable option guidance algorithms have been developed using reinforcement learning and a neural-based thermal locator to detect and utilize thermals[1]. The results showed that both heuristic controllers and reinforcement learning could be effectively combined with a thermal locating algorithm to improve UAV performance. Algorithms utilising reinforcement learning have proved too computationally expensive for real-world application, leaving robust but heuristic algorithms the only option with current processing power. These heuristic algorithms have been successfully employed [2][3] but there still is a desire to further optimise the aircraft trajectories.

For progress to be made in the improvement of the methods used to extract energy from atmospheric convection it is advantageous to start in a simulation environment before moving on to real-world flight tests. The use of a simulator allows the algorithms to be tested in a controlled environment where the conditions are both fully understood and repeatable. However for the results of the simulation to be both meaningful and useful the simulation environment must be realistic. Three key areas of the simulation environment need to accurately reflect reality; the aircraft flight dynamics, the atmospheric model and the aircraft flight control structure.

The following three sections deal with the aircraft flight dynamics, the atmospheric model and the aircraft flight control structure respectively.

Having established the simulation environment, section five and six shows how suitable atmospheric convection can be identified and exploited. Section seven gives some pertinent results. Section eight highlights the key conclusions and recommendations for real-world implementation.

## II. AIRCRAFT DYNAMICS MODEL

Autonomous soaring can be simulated with rudimentary knowledge of the aircraft in question but in order to optimise the algorithms an accurate model of the aircraft in question is required. The type of aircraft is unimportant for the purpose of optimisation as long as the actual aircraft is reflected. The X-Models XCalibur was chosen as the test aircraft as it has the best performance of all the aircraft available at Loughborough University. The X-Models XCalibur is a self launch electric glider with a 3.2m span and a typical take-off weight of 3.5kg, giving a wing loading of approximately 120Pa. The XCalibur was developed from X-Models F3J competition aircraft, and as such has a performance comparable with larger gliders used by other researchers[2][3]. An accurate model of the aircraft in use also facilitates the stabilisation and control algorithms to be validated in conjunction with the high-level trajectory generation algorithms.

It was decided to have a non-linear model that included the stall behaviour of the aircraft. This is important because of the possibility that the combination of the speed to fly and path planning algorithms might cause the aircraft to fly close to the stall condition. If the aircraft did stall, it is important to know how the control structure would behave.

The chosen environment for constructing the dynamics model was Matlab / Simulink. The dynamics model is made up of four distinct parts; left and right wings, elevator and rudder. The fuselage is neglected in the calculations because of its small influence in relation to the aircraft's responses and the difficulty involved in modelling it accurately. The glider dynamics model also includes a model of the motor, providing data for the simulation of the initial launch procedure and any subsequent powered flight that may be required.

The coefficients used in the model presented are based on the performance parameters of the XCalibur used at Loughborough University, as confirmed by flight tests.

## III. SHALLOW LAYER CUMULUS CONVECTION

In order to simulate and ultimately attempt to optimise the aircraft's trajectory, it is necessary to have a model of the atmospheric structures that the aircraft is flying through. Although ideally a model of high fidelity should be used to simulate and optimise the control algorithms, in reality a relatively crude model can be used effectively, provided that it can reflect the salient characteristics of the updraught structures.

Once the vertical motion of shallow layer convection is sufficiently strong to support the continued flight of an aircraft it is referred to as 'a thermal'. Nearly all of the existing models of thermals have been produced for hotter countries than the UK.

### A. Thermal Profiles

The 'British Standard Thermal' (BST)[4] that is used by the British Gliding Association (BGA) is given in the international system of units (SI) in (1).

$$w = 2.16[1 - (r/304.8)^2]$$

The magnitudes used in the British standard thermal (BST) will be used as a starting point to base the models on. It is worth stating that the BST is not a mean thermal for Britain but is an optimistic case that the BGA uses to assess full-size glider performance. The mean thermal strength that M. J. Allen[5] detected at Desert Rock in July of 2002 was 2.69m/s with a maximum strength of 6.3m/s. It is therefore suggested that the BST may represent a typical thermal on a summer's day in Britain. Although the BST is a good starting point, it does not model atmospheric sink or provide any information on the variation of strength or the radius of the thermal with height. Nor does the BST provide any information about the time dependent nature of the thermal. Once sufficiently centred in a thermal this detail has a negligible effect. However, when considering methods of centring it is advantageous to consider the foregoing factors.

It has long been known that downdraughts or associated with shallow layer cumulus convection [6][7]. M. J. Allen chose to largely ignore the associated downdraughts and arrived at a family of parametric profiles, as shown in *Fig. 1*[5]. Another parametric profile that has been used is that of J. Wharington [1], shown in *Fig. 2*. Although this model is less accurate than others [2][8], it lends itself well to mathematical analysis[3].

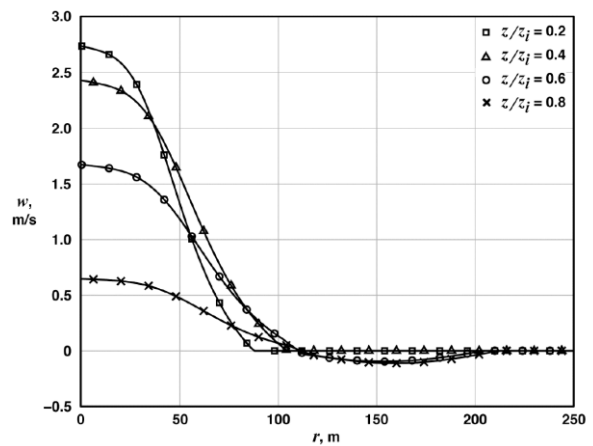


Figure 1. M.J. Allen – Vertical Thermal Velocity Prediction at height ratio of 0.4 [5]

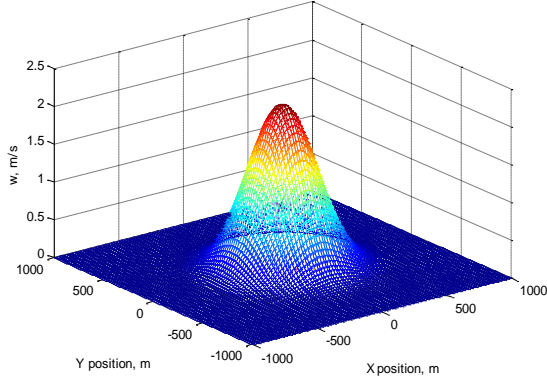


Figure 2. Thermal Lift Distribution as used by J. Wharington [1] [3]

None of the existing thermal distributions found meet what is felt to be a suitable starting distribution. The distributions either have spurious maxima or unrealistic sink associated with them. Allen's model was the best found but that did not, for the most part, include sink. Allen [5] used atmospheric data to generate the core velocities but his prediction of sink does not match anecdotal evidence of the almost inevitable presence of sink around the thermal as described in [7], [9] and [10]. The sink shown in Fig. 3 is exaggerated for illustrative purposes.

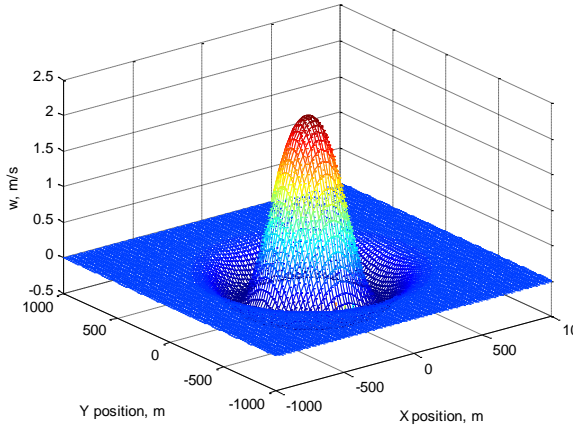


Figure 3. Proposed British Thermal Lift Distribution

Most models do not have sink associated with them but instead rely on conservation of mass to determine atmospheric sink. Although there is no denying that conservation of mass does apply to the global atmosphere, it may be argued that local weather systems will also have an effect. It can be frequently observed that regions of high pressure effectively suppress thermal formation over large areas of the country. Additionally, wave conditions and anticyclones can both suppress thermal formation. It is also known that strong down drafts form separately to known thermals. Conservation of mass is a good starting point for the atmospheric map but for large maps, overlying a large period sinusoidal distribution for atmospheric wave and 'cloud street' formation may yield improved realism [7][10][6].

As all the models looked at fall short in one way or another, a new thermal velocity distribution model is proposed as shown in Fig. 3. The model derived follows profile as measured by M. J. Allen [5] but using the generally accepted association of sink to a British thermal and is given in (2) below.

$$w = w_{max} \left( -C_1 \left( \frac{r}{\sqrt{C_1} r_{inner}} \right)^2 + 1 \right) e^{\left( -\left( \frac{r}{\sqrt{C_1} r_{inner}} \right)^{C_2} \right)}$$

$C_1$  and  $C_2$  control the radius and magnitude of the sink associated with the thermal structure.

### B. Thermal Spacing

Lenschow [11] derived an equation capable of estimating the distances between thermals at a constant height ratio,  $\frac{z}{z_i}$ , of 0.4. Where,  $z_i$  is the height of the convective layer and  $z$  is height. A guide to the distances between the thermals was given as 1.5 to 2.5 times the convective scale by C. E. Wallington in Meteorology for Glider Pilots [9]. John Delafield [6] suggested that the distances were between 2 and 3 times the convective scale in his book 'Gliding Competitively'. This would tend to suggest that the thermal spacing to height ratio is not consistent between different climates. An explanation of this phenomenon may be that there is a minimum spacing for the formation of thermals for them not to merge. In warmer climates the characteristic convective length scale will be much more than that of a temperate climate. The reduction in spacing of thermals in a temperate climate, although less than in warmer climates, is not sufficient to maintain an equivalent, thermal spacing to convective length scale ratio. Following the anecdotal evidence from [9] and [6], Lenschow's equations can be reworked using the numbers proposed above to take the following forms for a British climate.

$$\frac{N z_i}{L} = 0.5$$

$N$  is the number of updrafts encountered over a length  $L$ . This equation can be rearranged to give the number of updrafts in a given area.

$$N = \frac{0.25XY}{z_i r_{outer}}$$

Where,  $r_{outer}$  is the outer diameter of the average thermal.  $X$  and  $Y$  are the length and breadth of the area of interest. Although there is no way of verifying (3) and (4) without more accurate flight data, the predictions are more typical of the British weather pattern.

The positioning of the thermals on a given map is often given as random. On occasions when the thermals have no obvious trigger this is a fair assumption. However, in the British Isles the thermals often do have trigger points; a dark field, a power station, a factory, a motorway etc. The effect of trigger points is to set up streets of thermals up and down wind from the last.

Knowledge of this phenomenon can be used to aid the autonomous decision making processes.

#### IV. CONTROL STRUCTURE

The flight control system is critical to the successful execution of any generated trajectory. While the limitations of the flight control system will always impose constraints on the trajectory generation algorithms, it is advantageous to maximise the flight control performance to keep these practical limitations to a minimum.

For the trajectories generated to be useful it is imperative that the flight control system be representative and have comparable performance to the real platform. The flight control system is based on a nested PID architecture; the simulation was setup to reflect this.

The controller features a novel use of feed forward control as part of the pitch and yaw controllers. This is to pre-emptively suppress fluctuations in pitch and yaw; due to uncoordinated turns and adverse yaw. As these parts of the controller are a simple form of model predictive control they are more platform dependent than the other parameters.

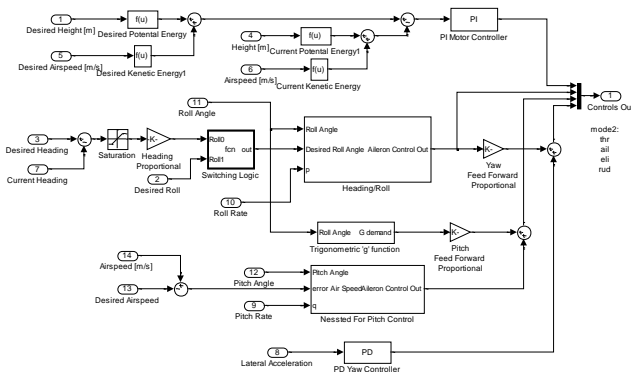


Figure 4. Control Structure

This control structure is not designed to reflect an optimal solution but instead accurately reflect the performance capabilities of the aircraft in question.

#### V. CHOICE OF AND EVALUATION OF THERMALS

The identification of suitable thermals has been investigated in detail over the years[3][7][9]. Equations to predicting the optimum speed to fly to optimise overall cross country speed are well known [4][12] but are incomplete without an estimate of the strength of the next thermal to be encountered; which is of course unknown until it is encountered. This prediction is generally referred to as the 'MacCready setting' after the first person to pose this problem. The choice of MacCready setting is a frequent topic of conversation at gliding clubs, but the problem boils down to how much risk can be tolerated. As the setting is a function of risk it follows that the setting is related to height, as a higher aircraft has a greater probability of encountering another thermal with the associated reduction in the risk of a forced landing or the use of a powered climb. D. Edwards [3] viewed landing out as unacceptable and his MacCready function

reflects this. Others have a higher tolerance to risk and as such select a more aggressive MacCready function. The XCalibur is fitted with a powerful electric motor so if the mission demanded maximum cross country speed at all cost, the MacCready function could be set aggressively resulting in profile that would ignore all but the strongest thermals, necessitating the use of the motor periodically. This powered climb and glide profile would be extremely power hungry, reducing range and crippling endurance. The choice of the function ultimately depends on the aircraft in question and the mission profile with the associated constraints. A good example of a constraint would be a maximum allowable height during the flight. The effect of this constraint would be to cause the MacCready function to tend to infinity at that height. In reality there are tolerances and safety margins. This type of constraint is shown in Fig. 5.

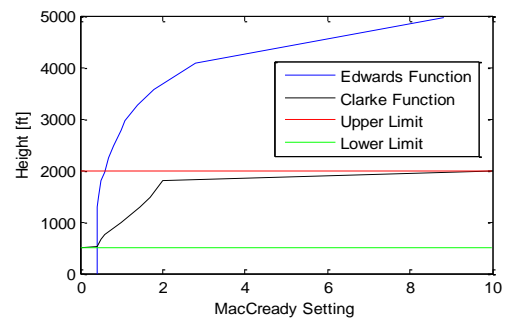


Figure 5. MacCready Setting

Once the MacCready setting has been established the aircraft can fly at the appropriate speed for the conditions and assess any thermals encountered for suitability. To facilitate the correct identification of thermals two variometer readings are used; an instantaneous reading and an averaged reading. The averaged reading suppresses the influence of turbulence and helps to prevent the erroneous thermal detection. Once a thermal stronger than the MacCready setting is detected the soaring algorithms are triggered. At that stage the averaged reading must drop below a lower critical value before the search for lift is abandoned. This is necessary because the aircraft will take a few turns to find the core of the thermal, with the sink that exists around the edge the average reading may fall considerably before stabilising. The instantaneous reading is used to position the aircraft in the thermal. Once the average reading has stabilised a decision can be made on whether the thermal is strong than the MacCready setting, if not the thermal is left in the hope of finding a stronger thermal along track. Similarly final glide calculations can also affect the MacCready setting as shown in Fig. 5, although this will not be considered further.

#### VI. CENTRING WITHIN A THERMAL

As time and height invariant thermal models aid visualisation, these have been used to illustrate the soaring techniques. However, accurate atmospheric models are required to validate the proposed techniques.

A thermal can be viewed as a vortex ring travelling upwards through the atmosphere with the aircraft's objective to be carried aloft in said vortex. In order to maximise the potential height gain of a given thermal the aircraft has to centre in the thermal as quickly as possible. If the aircraft does not find equilibrium inside the core of the thermal, then the aircraft will drop out of the bottom of the thermal. The factors that affect the aircraft's ability to find equilibrium include; the up draught strength, size, or inability to locate the strongest lift.

The inclusion of the associated 'sink' around the edge of a thermal is often neglected [1][2][3] because the sink found around the edge of very strong thermals is relatively small. However, in colder climates where the rise rate of the thermal may be lower compared to the vorticity of the thermal the sink around the thermal may be considerable. Once the aircraft is sufficiently well centred in the thermal the presence of sink around the edge of the thermal may be ignored but in order to evaluate the ability of a given algorithm to efficiently centre on a thermal the sink has a profound effect on the success rate.

There are many methods for centring in a thermal but two of the most widely used are the Piggot and the Reichman techniques.

"The point mass model simulation earlier demonstrates that Piggot's technique works well for negligible lag times and with perfect knowledge of the air mass velocity around the vehicle. For the full simulation model however, it appears that despite using accelerometers the response time is sufficiently long for Reichmann's technique to be more applicable than that of Piggot." [13]

It is therefore logical to base further work on Reichmann's technique. As discussed earlier, the time taken to centre in the core of the thermal is critical to the successful exploitation of the thermal encountered. As a result there is a desire to both better understand and to further optimise the positioning algorithms.

Ensuring the aircraft always turns in one direction while soaring allows the operator on the ground to quickly assess the flight mode the autopilot is currently in. It was shown by Piggott [7] that reversing the direction of the turn in a thermal is un-advisable, a turn direction monitor was added so that once a turn direction was chosen, it was not reversed.

Although there are algorithms to detect the relative location of the thermal with respect to the aircraft they are not infallible. This leads to a worst-case scenario of the aircraft turning in the wrong direction once encountering the edge of a thermal. This is the scenario that will be investigated when considering the stability of thermal location algorithms.

The control implemented in the simulation presented is a relatively simple implementation of the Reichman

method. Loughborough University operates a range of advanced autopilots. The autopilot that is fitted in the XCalibur is capable of accepting bank angle commands. This facilitates a more straightforward implementation of the Reichmann method.

Although the Reichman method provides good results, M. J. Allen [2] showed improvements by adding a thermal position estimator. The soaring controller therefore takes the following form shown in Fig. 7.

$$\begin{bmatrix} X_{Est} & Y_{Est} \end{bmatrix} = \begin{bmatrix} \left\{ \frac{(Vario)(X_{Position})}{Ks + 1} \right\} & \left\{ \frac{(Vario)(Y_{Position})}{Ks + 1} \right\} \\ \left\{ \frac{(Vario)}{Ks + 1} \right\} & \left\{ \frac{(Vario)}{Ks + 1} \right\} \end{bmatrix}$$

Although the (5) is in continuous time form it can be readily discretised for implementation with the real autopilot. This form of prediction has the advantage of having a variable sample length, for the prediction of the location of the thermal. The filter gain, K, was chosen as the time to complete one soaring turn, but this does not have to be the case.

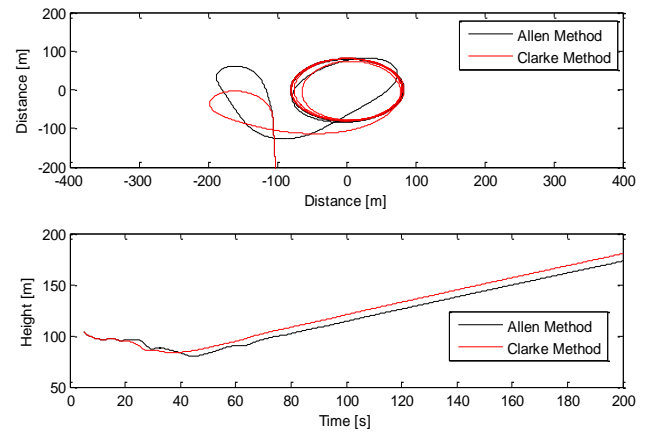
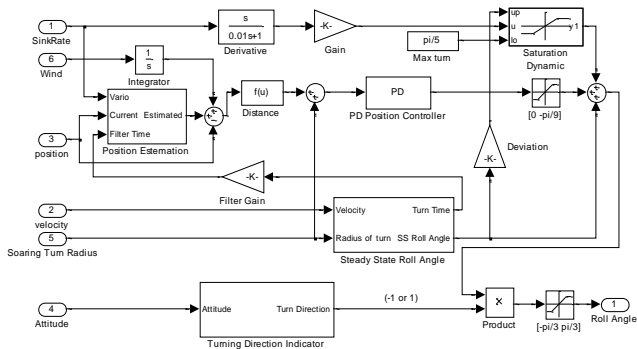


Figure 6. Performance Comparison

Although the Reichman-PD soaring controller shows improvements over M. J. Allen's method in the example quoted this is not always the case. This controller is of the same basic form as that used by M. J. Allen [2], with the exception that the autopilot he used could not accept bank angle commands, as a result his controller demands a turn rate. All that can be conclusively ascertained from the results is that the two methods are approximately equal. The advantage of the Reichman-PD soaring controller is that all of the terms in the controller have physical significance and are readily tuneable in real-time. Although the use of lowpass filters is more computationally intensive than other methods [1][2][3] it allows for dynamic adjustment of the number of datapoints used to predict the centre of the thermal.





To maximise the aircraft's chances of finding thermals and thus maximising its cross-country speed potential, the flight control algorithms utilise cloud street phenomena. To do this once a thermal has been found and utilised the trajectory is modified to fly directly into or downwind as long as this does not take the aircraft more than  $90^\circ$  off-track. The soaring controller also includes a prediction of the likely next thermal location along the current cloud street. If the location of this thermal would take it more  $90^\circ$  off-track the cloud street is also rejected. This projection is based on the convective scale assumptions presented in section III b.

It can be seen that the aircraft proceeds to take five thermals along the first cloud street before abandoning to continue on track. The aircraft fails to contact another cloud street so the motor is used towards the end of the flight to ensure the goal is reached. Having used the

motor to gain altitude, another thermal is contacted allowing more energy to be harvested before the task is completed. Although the motor was used, an energy-saving of more than 90% was demonstrated when compared to the same task being completed under conventional powered flight by the same aircraft.

## VIII. SUMMARY AND CONCLUDING REMARKS

A suitable simulation model has been developed consisting of a non-linear aircraft dynamics model, a parametric thermal model and a realistic control structure. The parametric thermal model was updated from those previously used to more accurately reflect the British climate. The presence of nontrivial amounts of sink associated with the thermal structure along with the prevalence of cloud streets has been reflected in the atmospheric model. The practical implementation of the MacCready function with restrictive height constraints has been discussed and implemented. A new flexible implementation of the Reichman centring technique was proposed and evaluated, providing promising results. These disparate elements were finally brought together in a simulated task. The simulated task was a 10 km outbound journey in challenging conditions. Despite the aircraft not completing the task without using its motor an energy-saving of more than 90% was demonstrated when compared to the same flight completed without the use of thermals.

## REFERENCES

- [1] J. Hersington, and I. Herszberg, “Control of a High Endurance Unmanned Air Vehicle,” ICAS-98-3,7,1, AIAA A98-31555, 21st ICAS Congress, Melbourne, Australia, September 13–18, 1998.
- [2] M. J. Allen, “Autonomous Soaring for Improved Endurance of a Small Uninhabited Air Vehicle.” NASA Dryden Flight Research Center, Edwards, California, 2005.
- [3] D. J. Edwards, “Implementation Details and Flight Test Results of an Autonomous Soaring Controller”. s.l.: North Carolina State University, Raleigh, NC, 27604, 2008.
- [4] F. Irving, “The Paths of Soaring Flight”. London: Imperial College press, 1999, 2006. ISBN 1-86094-055-2
- [5] M. J. Allen, “Updraft Model for Development of Autonomous Soaring Uninhabited Air Vehicles” NASA Dryden Research Centre, Edwards, California, USA 2005
- [6] J. A. Kyle, “Optimal Soaring by a Small Autonomous Glider”. Oregon: Oregon State University, 2006.
- [7] D. Piggott, “Gliding, A Handbook on Soaring Flight”. London: A & C Black, 2002, 1997, 1990, 1986, 1976, 1971, 1967, 1958. ISBN 0 7136 6148 8.
- [8] C. E. Wallington, “Meteorology for Glider Pilots”, Third International Edition, 1989 ISBN 0-7195-3303-1
- [9] A. Welch, “Pilots’ Whether”. London: William Clowes and sons, Ltd, 1979, 1973. ISBN 0 7195 2661 2.
- [10] J. Delafield, “Gliding Competitively, 1982”, ISBN 0-7136-2224-5
- [11] D. H. Lenschow, and P. L. Stephens, “The Role of Thermals in the Convective Boundary Layer”. s.l.: Boundary-Layer Meteorology, 1980
- [12] P. B. MacCready, Jr, “Optimum airspeed selector”, Soaring, pp. 10-11. 1958
- [13] I. D. Cowling, S. Wilcox, Y. Patel, P. Smith. “Increasing persistence of UAVs and MAVs through thermal soaring”. 1145. Clapham: The Aeronautical Journal, 2009. Vol

# Multi-objective Optimization of Constrained Parallel Hybrid Electric Vehicles

Shaobo Li, Jinglei Qu

Key Laboratory of Advanced Manufacturing Technology (Guizhou  
University)  
Ministry of Education  
Guiyang, China

Guanci Yang

Chengdu Institute of Computer Applications  
Chinese Academy of Sciences  
Chengdu, China  
guanci\_yang@163.com

**Abstract**—Hybrid Electric Vehicles (HEVs), surrounded by high complexity, nonlinear constraint and large amount of coupling design parameters, provides fairly higher fuel economy with lower emissions than conventional vehicles. It is significant to optimize HEV's parameters to enhance its performance. Considering the disadvantage of the methods transforming multi-objective functions into a single objective evaluation function, this paper reports a methodological approach for multi-objective optimization of parallel hybrid vehicle. Firstly, a model of parallel hybrid electric vehicle for optimal simulation is established. Secondly, based on the non-dominated sorting genetic algorithms II, a methodological approach for the simultaneous optimization of HEV parameters to minimize the fuel consumption and emissions was proposed, which adopts ADVISOR to simulate. Taking Insight as a case, the simulation results show that this approach can obtain a set of Pareto-optimal solutions with better performance.

**Keywords**—constrained multi-objective optimization; hybrid system; multi-objective evolutionary algorithm

## I. INTRODUCTION

Hybrid Electric Vehicles (HEVs) are a hotspot of automobile industry and research for their beneficial effect on environment and energy consumption, which can improve the fuel efficiency and reduce the emissions by the use of power control strategy. The goal of power control is to satisfy power requirement with the cooperation of electric power source and fuel power source. Synchronously, the corresponding fuel consumption and emissions are as low as possible.

The HEV design problem usually aims at several simultaneous objectives. The primary goal is the minimization of the vehicle fuel consumption and emissions in the condition of maintaining or enhancing driving performance. Gradient-based algorithms, such as Sequential Quadratic Programming (SQP), use the derivative information to solve this problem [1, 2]. The major disadvantage of these methods is that they are weak at global optimization. Meanwhile, these search techniques require strong assumptions for the objective function, such as continuity, differentiability, satisfaction of the Lipschitz condition etc., which cannot be trivially assumed for this problem. Reference [3] uses Genetic Algorithms (GAs) to find out the optimal component sizes in HEV. Its objective is to minimize a weighted sum of fuel consumption and emissions, and the dynamic performance of the vehicle is applied as constraints.

Reference [4] applies GA to simultaneous optimal parallel HEV component and control strategy. The parallel HEV together with the Electric Assist Control Strategy (EACS) are employed to formulate the optimization problem. That it constructs a function by allocating weights to each of the objective for the initial multi-objective problem is transformed into a mono-objective problem, which encompasses fuel consumption and exhaust emissions such as CO, HC and NO<sub>x</sub>, and the PNGV performance requirements including acceleration and gradeability characteristics are considered as constraints. Those researches have a common feature that many multi-objective optimizations of hybrid system employ single objective optimization methods by transforming multi-objective functions into a single objective evaluation function. In order to eliminate weighting coefficient to reflect the nature of each objective, this paper proposes a kind of method that constrained parallel hybrid system optimization based on multi-objective evolutionary algorithm.

## II. MODEL OF PARALLEL HYBRID ELECTRIC VEHICLE

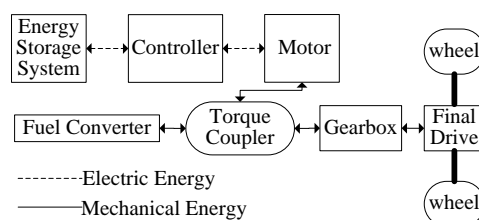


Figure 1. Parallel hybrid electric vehicle structure

Parallel hybrid electric vehicle (PHEV) drive system includes an engine with an associated electric motor as shown in Fig. 1, both engine and electric motor can deliver power to wheels, and the engine is the main drive role [5]. Under an intermediate-speed or fast-speed driving, engine will be the main drive mode, otherwise, the motor will play a leading role especially driving with low-speed or big torque, the electric motor also may be used as a generator to charge the battery by either the regenerative braking or absorbing the excess power from the engine when its output is greater than that required to drive the wheels [6]. In this working mode, the electric motor has low-speed with an export high torque, which can make the motor play a full function and let the engine work in more efficient region so as to avoid some adverse

woke conditions, such as starting, accelerating or climbing, to decrease fuel consumption and emissions. Because of that PHEV has two power producers, and it requires an engine and an electric motor with smaller size and lighter weight to provide the same performance comparing with the traditional drive system, which makes the PHEV more suitable for passenger cars [7]. Although the performance of PHEV is closed to traditional cars, it has more obvious advantages in fuel consumption and emissions.

### III. CONTROL STRATEGIES

In general, power control strategies can be roughly classified into three kinds:

1) Rule-based algorithms, such as energy following and thermostatic.

2) Real-time optimization. Several algorithms have been proposed for real-time optimization, including fuzzy logic controller [8] and energy-flow analysis. The control strategy with real-time optimization calculates the optimal torque based on the engine feature parameters, and decides the actual torque output by modifying the optimal torque based on real-time road situation and battery State Of Charge (SOC). Dynamic optimization parameters can be changed based on real-time operation state and energy requirement, then present a real-time optimal solution for power control. But the heavy computation and real-time requirement are the challenges of real-time optimization.

3) Static optimization method, such as dynamic programming, sequential quadratic programming, baseline strategy, DIRECT, as so on [9]. The optimization solutions figure out the proper split between the electric power and fuel power using steady-state efficiency maps, and realize HEV optimization based on the operation with the optimal parameters [10].

Logic threshold control method is a typical rule-based algorithm. It is characterized with quick operation, easy realization and high practicability, which is widely applied to the hybrid electric vehicles [11]. The EACS is a regular logic threshold control method, which uses the motor for additional power when needed by the vehicle and maintains charge in the batteries. This strategy has a set of static parameters to limit the working area of engine and some pre-set rules to judge or determine the working mode of hybrid system. The vehicle will be operated by the EACS methodology based on two rules which is illustrated in Fig. 2. Fig. 2-a) shows that the  $V_{SOC}$ , the value of battery state of charge, is higher than its low limit ( $L_{SOC}$ ), if the required speed is less than a given value which is called the electric launch speed, and then the engine will turn off. Furthermore, if the required torque is less than a cutoff torque that is referred to as ‘Engine OFF’ fraction, the engine will also turn off. Fig. 2-b) illustrates the case when the  $V_{SOC}$  is lower than  $L_{SOC}$ . In this case 1, an additional torque is required from the engine to charge the battery. This additional charging torque is proportional to the difference between  $V_{SOC}$  and the average of  $H_{SOC}$  and  $L_{SOC}$ . Only when the engine is on, the Engine charging torque is required. The engine torque is prevented from being below a certain fraction of the

maximum engine torque that is referred to as ‘Engine On’ fraction. This strategy is designed to prevent the engine from operating at low torque condition.

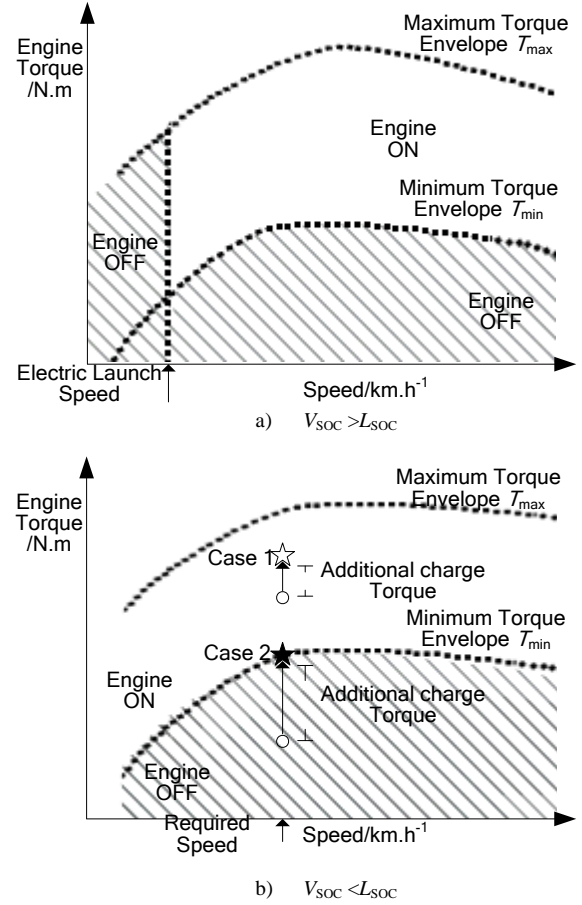


Figure 2. Influence of the control strategy on the engine operation. Operation point requires plus charge torque in case 1, and operation point lies along minimum torque envelope because required torque plus charge torque is too low in case 2.

### IV. MULTI-OBJECTIVE OPTIMIZATION MODEL OF CONSTRAINED HYBRID SYSTEM

#### A. Mathematical Description

Essentially, PHEVs optimization is a Multi-objective Problems (MOPs) which have two major objectives (improve efficiency of fuel and reduce emission) under constraint (good drive performance). The mathematical description of such multi-objective optimization problems is as follows [12]:

$$\begin{cases} \min & \mathbf{y} = F(\mathbf{x}) \Delta (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))^T \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0 (i=1, 2, \dots, q_1) \\ & h_j(\mathbf{x}) = 0 (j=1, 2, \dots, q_2) \end{cases} \quad (1)$$

Where  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in X \subset R^n$  is the variable vector,  $\mathbf{y} = (y_1, y_2, \dots, y_m) \in Y \subset R^m$  is the feasible solution space,  $F(\mathbf{x})$  is defined as  $m$  mapping functions, namely,  $f: X \rightarrow Y$ ,  $g_i(\mathbf{x}) \geq 0 (i=1, 2, \dots, q_1)$  and  $h_j(\mathbf{x}) = 0 (j=1, 2, \dots, q_2)$  are the given criteria. In this study,

the fuel consumption and the emissions of various pollutants are employed to formulate the optimization problem, and the objective functions are defined as follows:

$$f_1(\mathbf{x}) = \text{Fuel}(\mathbf{x}) \quad (2)$$

$$f_2(\mathbf{x}) = \text{HC}(\mathbf{x}) + \text{NO}_x(\mathbf{x}) + 0.1\text{CO}(\mathbf{x}) \quad (3)$$

Where  $\text{HC}(\mathbf{x})$ ,  $\text{CO}(\mathbf{x})$  and  $\text{NO}_x(\mathbf{x})$  are the exhaust emissions of hydrocarbons, carbon monoxide and nitrogen oxides, which have the common unit of measure, and therefore summarized as one objective function. Note that the exhaust emission of carbon monoxide is about ten times than others, thus the coefficient of  $\text{CO}(\mathbf{x})$  is 0.1.

### B. Constraints of the Optimization Problem

The indicator of automobile dynamic quality mainly includes accelerating ability and gradeability [13]. The vehicle acceleration performance usually is represented by acceleration time which includes vehicle standing start acceleration time and overtaking acceleration time. Standing start acceleration time is a duration that from a vehicle starts at the first block to a predetermined speed. Overtaking acceleration time indicates the vehicle accelerate to a high speed with the highest or second highest gear. The largest degree of a vehicle climbing with a certain value in a given time is called gradeability. One hybrid electric system has two sets of energy systems, so the energy output by battery must be considered when we calculate the total fuel consumption, and it is a common solution to convert the energy supplied by battery into fuel consumption of engine. In this study, to obtain the specific automobile dynamic quality and SOC balance is the one goal of optimization. Some important constraints are shown in Table I.

TABLE I. CONSTRAINTS TO OPTIMIZE PHEVS

Item	Description	Condition
automobile dynamic quality	the acceleration time of 0-100km/h	$\leq 14\text{s}$
	the acceleration time of 40-100km/h	$\leq 10\text{s}$
	the climbing degree of continuous running 10s with 20km/h	$\geq 30$
SOC balance	the difference between the initial and final values of SOC during a certain driving cycle	$\leq 0.5\%$

### C. Optimization parameters

Considering the parameters that the maximum power of engine ( $P_e$ ), the modules number of battery ( $N_b$ ) and the final driver ratio of vehicle ( $R_m$ ) have greatly influence on the PHEV's maximum speed, acceleration and driving range, we select those parameters as the most vital power train factors that need to optimize, and the upper ( $H_{\text{SOC}}$ ) and lower ( $L_{\text{SOC}}$ ) bounds of SOC are chosen, which have a crucial influence over vehicle's power performance, emission and fuel economy. The parameters of PHEV and their range are listed in Table II.

TABLE II. PARAMETERS RANGE OF THE PHEV NEED TO OPTIMIZE

Name	Value
$P_e/\text{kw}$	[20,70]
$N_b$	{15,16,...,49,50}
$H_{\text{SOC}}$	[0.5,0.9]
$L_{\text{SOC}}$	[0.1,0.5]
$R_m$	[0.5,10]

## V. MULTI-OBJECTIVE OPTIMIZATION ALGORITHM FOR HYBRID SYSTEM

NSGA- II [14] was proposed by Deb et al, which is based on non-dominated sorting and elitism strategy. NSGA- II has excellent performance on computational complexity and robustness, and it is the most cited algorithm on the field of evolutionary multi-objective optimization by SCI. For convenience,  $t$  denotes the current generation, the  $t$ -th generation of the evolutionary population is  $P_t$ ,  $Q_t$  is used to store new individuals of the  $t$ -th generation,  $R_t$  is the temporary mating population,  $N$  is population size of  $P_t$  and  $G$  is maximum evolution generation. The multi-objective optimization algorithm of hybrid electric system based on NSGA- II (HES- NSGA- II) is illustrated as Fig. 3.

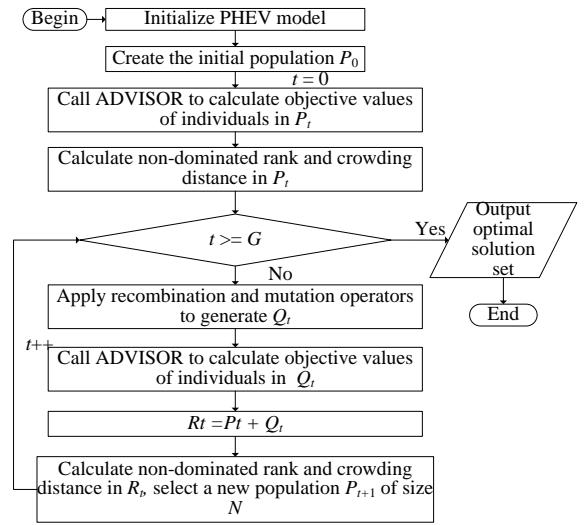


Figure 3. The flow of multi-objective optimization algorithm of hybrid system based on NSGA- II

## VI. SIMULATION AND ANALYSIS

### A. Parameters

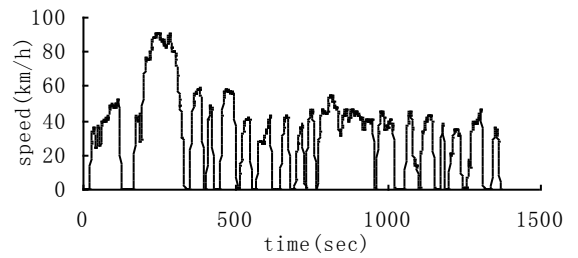


Figure 4. UDDS driving cycle

The algorithm designed in section V was implemented in Matlab6.5. Each operator is subject to NSGA- II. Table II shows the variables and their range. Setting the evolution population size is 50, and the maximum evolution generation is 100, and the mutation rate is 0.1, crossover probability is 0.9. We use ADVISOR to simulate hybrid electric vehicle and the Insight is instantiated as the test PHEV. Because of that ADVISOR had not been model the pollutant data module for the engine in Insight, we have modified and improved it by adding required

module refer to a similar-sized gasoline engine model (see FC\_SI41\_emis.M), and its engine displacement is 1.0L and the maximum power is 41kw. The urban dynamometer driving schedule (UDDS) mode driving cycle is used for the optimal simulation, and the speed profile of the UDDS driving cycle is shown in Fig. 4. The duration of UDDS driving cycle is 1369s, and the distance of the whole driving cycle is 11.99km, and the maximum speed is 91.25km/h. More detailed characteristics of UDDS see Table III.

TABLE III. CHARACTERISTIC OF UDDS

Item	Value	Item	Value
Simulation time	1369s	Average acceleration	0.5 m/s <sup>2</sup>
Distance	11.99km	Maximum deceleration	-1.48 m/s <sup>2</sup>
Maximum speed	91.25km/h	Average deceleration	-0.58 m/s <sup>2</sup>
Average speed	31.51km/h	Idle time	259s
Maximum acceleration	1.48m/s <sup>2</sup>		

### B. Results and Analysis

According to the settings given in section A, we have run HES- NSGA- II 10 times independently. With the preference of minimize fuel consumption and emissions, the selected Pareto optimal solutions are shown in Table IV, in which No.1 data is Insight's performance on fuel consumption and emissions with the default settings in ADVISOR[1], and the others are the optimized parameters and objectives. To observe Table IV, it is obvious that the optimized objectives are lower than before, which demonstrates that HES-NSGA- II can improves the performance of Insight.

TABLE IV. THE SELECTED OPTIMAL SOLUTIONS FORM 10 INDEPENDENT RUNNING

No.	$L_{soc}$	$H_{soc}$	$P_e$	$R_m$	$N_b$	$f_1$	$f_2$	HC	NO	CO
1	0.20	0.80	49.9205	1.00	20	3.8731	0.7960	0.3600	0.2620	1.7420
2	0.36	0.62	42.7577	1.16	49	3.8557	0.7791	0.3601	0.2509	1.6814
3	0.37	0.76	55.5996	1.16	38	3.8576	0.7776	0.3595	0.2500	1.6805
4	0.45	0.70	34.5195	1.17	31	3.8596	0.7749	0.3584	0.2493	1.6713
5	0.36	0.62	44.9862	1.17	49	3.8624	0.7743	0.3582	0.2490	1.6708
6	0.40	0.57	42.1202	1.18	49	3.8644	0.7736	0.3579	0.2486	1.6703
7	0.35	0.62	43.0832	1.18	50	3.8656	0.7731	0.3578	0.2484	1.6696
8	0.40	0.55	48.6172	1.19	31	3.8674	0.7730	0.3581	0.2480	1.6690
9	0.35	0.61	44.3202	1.19	50	3.8686	0.7718	0.3576	0.2474	1.6682
10	0.30	0.70	64.2809	1.19	43	3.8698	0.7717	0.3575	0.2473	1.6682
11	0.24	0.64	34.9336	1.20	32	3.8712	0.7695	0.3566	0.2465	1.6643

Taking an optimal result as a case from the 10 runs randomly, and the distribution of the Pareto optimal solutions is shown in Fig. 5. From this figure it can be found out that using HES-NSGA- II to optimal PHEV parameters can obtain a set of alternative solutions.

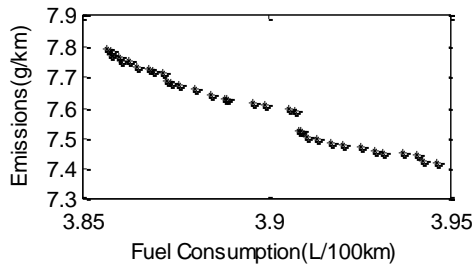


Figure 5. Distribution of Pareto optimal solutions of one run

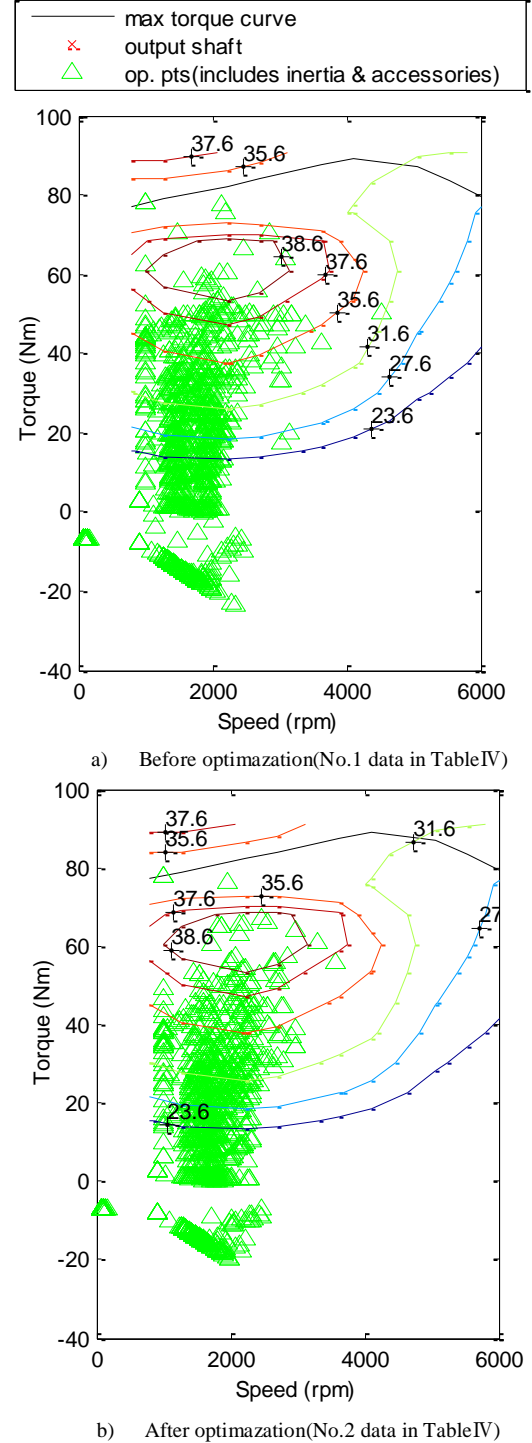


Figure 6. Engine operation

Let the No.2 data is the setting of Insight, and its engine operation is detailed in Fig. 6. Comparing a) with b) in Fig. 6, it illustrates that the engine's working states are more concentrated than before, which means that the engine optimized by HES- NSGA- II has better performance on matching power transmission system and is more efficiency. Note that the control strategy adopted by Insight does not restrict the working fraction of engine, so the engine will still work on a low torque condition.

The fuel converter efficiency is detailed in Fig. 7-b) and its motor/controller efficiency is illustrated in Fig. 8-

b). Compared with default setting, the obtained solutions can satisfy the constraints better and reduce the fuel consumption as well as emissions simultaneously. Fig. 7 shows the compare of fuel converter efficiency between group 1(original setting of system) and 2(optimized setting of system). Though the figure we can get the information that the fuel converter efficiency mainly distribute in the range of [0.15, 0.35] in Fig. 7-a) and the other's in the range of [0.25, 0.35] (see Fig. 7-b)), which means that the fuel converter efficiency of engine is improved, which is benefit to improve fuel economy.

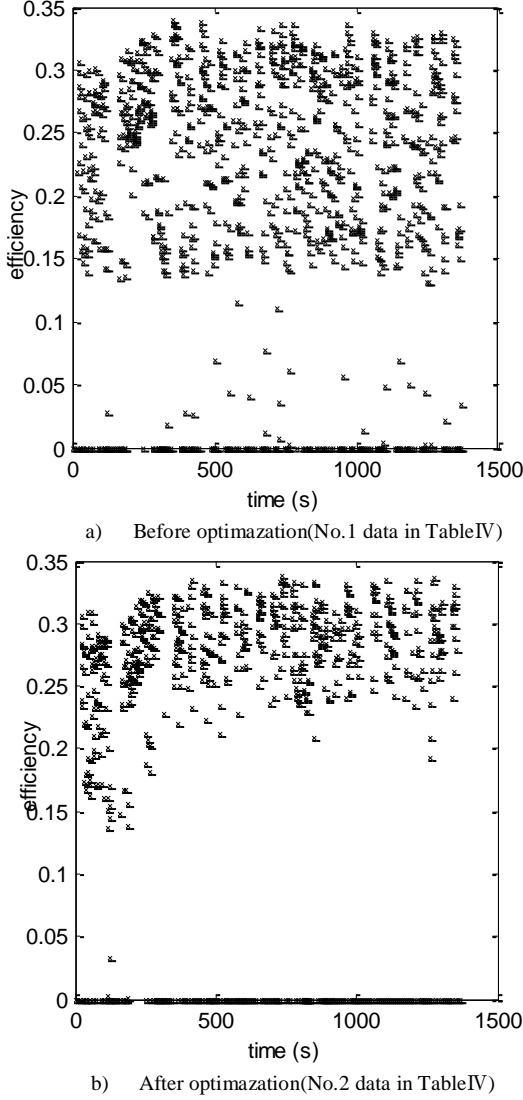


Figure 7. Compare of Fuel Converter Efficiency between group 1(default setting) and 2(optimized setting of system).

To observe Fig. 8, It is obvious that the motor/controller efficiency points locating in [0.2, 1] are increased considerably comparing with Fig. 8-a), and it suggests that the improvement of motor/controller efficiency is in favor of the decrease of fuel consumption and exhaust emissions.

From the above analysis, it can be point out that the HES-NSGA- II can enhance the efficiency of engine and motor, and has benefit of reducing the fuel consumption

and emissions of PHEV. Namely the HES-NSGA- II has advantages on optimizing the parameters of drive system and control strategy. In a word, HES-NSGA- II is capable to reduce the fuel consumption and emissions of PHEV and it also can provide a set of alternative Pareto-optimal solutions for user to satisfy the various requirements.

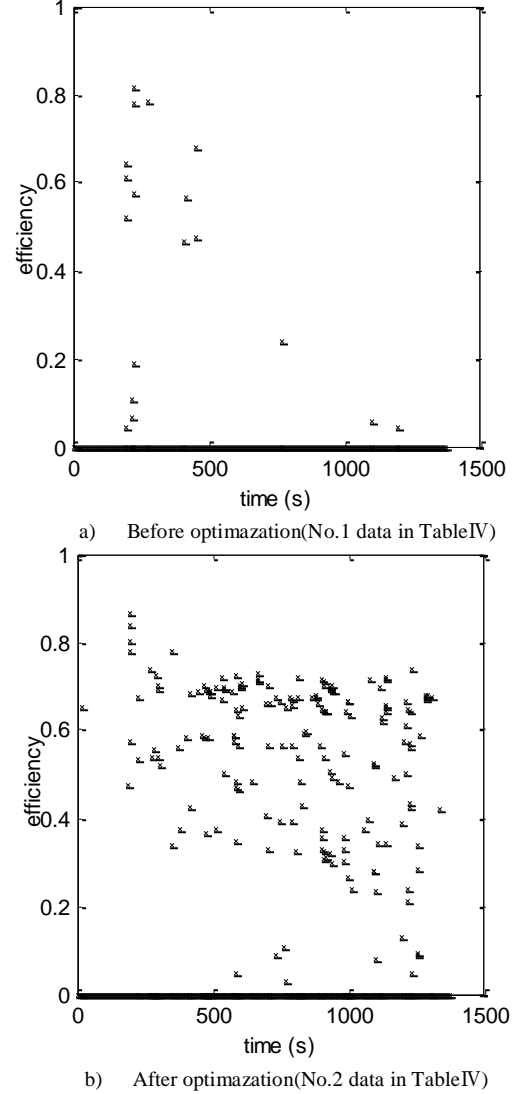


Figure 8. Compare of motor/controller efficiency between group 1(default setting) and 2(driving only, not regeneration).

## VII. CONCLUSIONS

Hybrid Electric Vehicles is a hotspot which represent future vehicle. Essentially, HEVs optimization is a Multi-objective problems which have two major objectives (improve efficiency of fuel and reduce emission) under constraint (good drive performance). A major goal of this paper is to provide a methodological approach, HES-NSGA- II, through the analysis of PHEV and NSGA- II for the simultaneous optimization of PHEV. From the results and analysis, we can make a conclusion that HES-NSGA- II can provides a series of Pareto-optimal solutions with higher fuel economy and lower emissions without sacrificing the performance of PHEV, which can provide a wide range of candidate solutions for powertrain and



control strategy variable parameters simultaneously to satisfy the various requirements. The work provides the guidance and a new way to optimize HEVs.

#### REFERENCES

- [1] K. Wipke, T. Markel, "Optimization techniques for hybrid electric vehicle analysis using ADVISOR," ASME International Mechanical Engineering Congress and Exposition, USA, p. 11–16, 2001.
- [2] R. Fellini, N. Michelena, P. Papalambros, M. Sasena, "Optimal design of automotive hybrid powertrain systems," First International Symposium on Environmentally Conscious Design and Inverse Manufacturing, Japan, p. 1–4, 1999.
- [3] M. Montazeri-Gh, A. Poursamad, "Optimization of component sizes in hybrid electric vehicle via genetic algorithms," ASME International Mechanical Engineering Congress and Exposition, USA, p.1-6, 2005.
- [4] M. Montazeri-Gh, A. Poursamad, "Application of genetic algorithm for simultaneous optimisation of HEV component sizing and control strategy," *Int. J. Alternative Propulsion*, vol. 1, p. 63-78, January 2006.
- [5] N. Schouten, M. Salman, N. Kheir, "Energy management strategies for parallel hybrid vehicles using fuzzy logic," *Control Engineering Practice*, vol. 11, p. 171-177, March 2003.
- [6] A. Sciarretta, M. Back, L. Guzzella, "Optimal control of parallel hybrid electric vehicles," *IEEE Trans. on Control Systems Technology*, vol. 12, p. 352-363, March 2004.
- [7] Q. Wang, P. Spronck, R. Tracht, "An overview of genetic algorithms applied to control engineering problems," *Second International Conference on Machine Learning and Cybernetics, China*, p. 651-1 656, 2003.
- [8] H D Lee, S K Sul, "Fuzzy-logic-based torque control strategy for parallel-type hybrid electric vehicle," *IEEE Trans. on Industrial Electronics*, vol. 45, p.625-632, August 1998.
- [9] P. Antonio, I. Lucio, Z. Vincen, V. Alfredo, "Optimization of energy flow management in hybrid electric vehicles via genetic algorithm," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Italy*, p. 1-5, July 2001.
- [10] X. Zhang, J. Song, Y. Tian, X. Zhang, "Multi-objective optimization of hybrid electric vehicle control strategy with genetic algorithm," *Chinese Journal of Mechanical Engineering*, vol. 45, p. 36-40, February 2009.
- [11] Z. Lian, Y. Deng, C. Yan, "Parameters optimization of a hybrid electric vehicle based on niche genetic algorithms," *Journal of Wuhan University of Technology*, vol. 31, p. 102-105, May 2009.
- [12] H. Ishibuchi, T. Murata, "multi-objective genetic local search algorithm and its application to flowshop scheduling," *IEEE Trans. Syst. Man Cybern.* vol. 28, p. 392–403, August 1998.
- [13] A. Emadi, M. Ehsani, J. Miller, *Vehicular electric power systems: land, sea, air and space vehicle*. NY: Marcel Dekker INC, 2003.
- [14] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Trans. on Evolutionary Computation*, vol. 6, p. 182-197, February 2002.

# Morphological Filters Based on Motif Combination for Functional Surface Evaluation

Shan Lou, Xiangqian Jiang, Paul J. Scott

Centre for Precision Technologies, University of Huddersfield  
Huddersfield, HD1 3DH, UK  
s.lou@hud.ac.uk

**Abstract**—Regarded as the complement of commonly used mean-line based filters, morphological filters are more function oriented. They are relevant to geometrical properties of surfaces and provide better results for functional evaluation of surfaces. The paper first gives a brief introduction to morphological filters. An algorithm to implement morphological filters is proposed, which is based on the motif combination. Experimental data are presented to illustrate the algorithm's superiority in performance. The end effect of morphological filters is corrected by the reflective padding. Either circular or horizontal structuring element is available using this method. Two examples of applying the morphological closing filter with the disk and the line-segment structuring element on a milled surface profile are illustrated. Finally, the morphological alternating sequential filter is employed to evaluate the roughness of functional stratified surfaces as a replacement to the two-stage Gaussian filter.

**Keywords**—morphological filters; motif combination; surface texture; functional analysis

## I. INTRODUCTION

Surface finished has always been important in engineering as it plays two critical roles: on one hand it helps to control the manufacturing process; On the other hand it helps functional prediction. Over the years, more attentions were given to the former, while less work has been done for functional evaluation. This leads to the abundance of mean-line based evaluating techniques (Gaussian filter, spline filter etc.) and the propagation of characterizing parameters (Ra, Rz, Rq etc.) [1]. However mean-line based techniques and parameters are designed to describe the average statistical characteristics of surface textures. In contrast, the envelope filter, achieved by rolling a ball over the surface, is more related with geometrical properties of surfaces and offers better results for functional prediction of surfaces [2].

With the introduction of mathematical morphology, morphological filters emerged as the evolution of the traditional envelope method [3, 4]. Morphological filters are essentially the superset of the early envelope filter, offering more tools and capabilities. They are carried out by performing morphological operations on the input surface with circular or flat structuring elements. Over the last decade, morphological filters have found many applications in practice. The morphological closing filter was utilized to approximate the conformable interface of two mating surfaces [5]. The morphological alternating symmetrical filter was employed to decompose the surface

topography of an internal combustion engine cylinder [6]. ISO 16610-49 [7] illustrated an example of detecting the defective milling mark from a milled surface using the morphological scale-space technique.

This paper proposes a morphological method based on the motif combination and illustrates its usage for the evaluation of stratified functional surfaces. Section 2 gives a brief introduction to morphological filters. An algorithm based on the motif combination is presented in Section 3. Section 4 illustrates two examples of applying morphological closing filters using the disk and the line-segment structuring element respectively. A discussion to the algorithm is given in Section 5. In section 6, we use morphological filters to evaluate stratified functional surfaces and demonstrate their superiorities over the two-stage Gaussian filter. Finally Section 7 reaches the conclusion.

## II. MORPHOLOGICAL FILTERS

Morphological filters are based on four basic morphological operations, namely dilation, erosion, opening and closing. They form the foundation of mathematical morphology [8].

Dilation combines two sets using the vector addition of set elements. The dilation of  $A$  by  $B$  is:

$$D(A, B) = A \oplus \overset{\vee}{B}.$$

where  $\overset{\vee}{B}$  is the reflection of  $B$  through the origin of  $B$ .

Erosion is the morphological dual to dilation. It combines two sets using the vector subtraction of set elements. The erosion of  $A$  by  $B$  is:

$$E(A, B) = A \ominus \overset{\vee}{B}.$$

Opening and closing are dilation and erosion combined pairs in sequence. The closing of  $A$  by  $B$  is given by applying the dilation followed by the erosion,

$$C(A, B) = E(D(A, B), \overset{\vee}{B}).$$

Fig. 1 demonstrates an example of the closing of an open profile by a disk. The closing envelope is the lower locus of the disk rolling over the measured profile from above.

Opening is the morphological dual to closing. The opening of  $A$  by  $B$  is given by applying the erosion followed by the dilation,

$$O(A, B) = D(E(A, B), B).$$

In contrast to Fig. 1, Fig. 2 presents the opening envelope of the profile which is obtained by rolling the disk over the profile from below.

Morphological operations are nothing new in surface texture analysis. The scanning of the tactile stylus over the workpieces surfaces, as a common practice in roughness measurement, is a morphological dilation operation. The mechanical surface could be reconstructed by carrying out on the measurement data the erosion operation with a sphere of the same radius as the stylus [9].

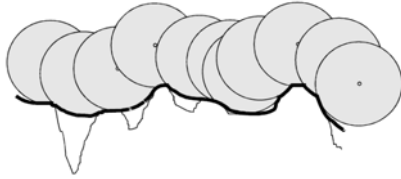


Figure 1. Closing of an open profile by a disk.

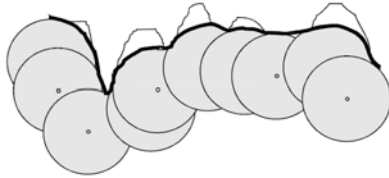


Figure 2. Opening of an open profile by a disk.

### III. ALGORITHM BASED ON MOTIF COMBINATION

ISO 16610-41 [10] presents a basic method to compute discrete morphological filters. It puts the origin of the structuring element at every point of the input profile, as illustrated for a few positions of a circular structuring element for dilation in Fig. 3. Extreme value at each position is collected and they form the output envelope. The extreme heights for input points are the results of adding the ordinates of input profile points with the ordinates of sampled points on the disk, as marked by the top-most stars at vertical lines in the figure. Due to the duality of the dilation and the erosion, the erosion could be easily computed by first flipping the structuring element and later flipping the dilation results,

$$E(S, B) = -D(S, -B).$$

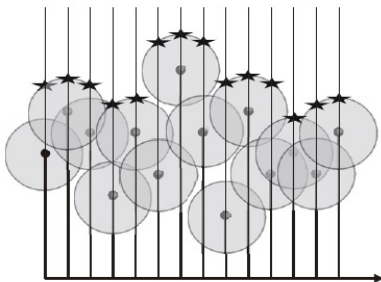


Figure 3. Dilation of discrete measured points with a disk.

This method follows the definition of morphological operations and therefore it is called as the naive algorithm. Scott [11] proposed an alternative way to calculate the profile envelope using the motif combination. A couple of definitions were given as the data type used in the motif combination algorithm.

Events: an event split the profile into a number of discrete sections. The events might be the highest points on all the local peaks or all the upcrossing of the profile through a reference line or even every sample point of the profile. They are numbered in order along the profile. The initial set of events is all the sample points on the profile.

Motif: a motif  $(i, j)$ , where  $i < j$ , consists of that section of the profile between the  $i$ th and  $j$ th events.

Motif Combination Test: it is performed on two adjacent motifs (say, two motifs  $(i, j)$  and  $(j, k)$ ) with the common event (say,  $j$ ) to determine if the common event is significant or not. If the event is not significant, two adjacent motifs to that event are combined (say, motifs  $(i, j)$  and  $(j, k)$  are combined to form a new motif  $(i, k)$ ) and thus the event is eliminated.

The motif combination procedure eliminates insignificant motifs and obtains significant ones. It is consistent with the functionality of morphological filter in that the features on the profile smaller than the structuring element in size are removed by the filters. This consistency provides an access to computing morphological filters by means of the motif combination. For rolling a disk on the profile, the functional motif combination test is to check if the disk is possible to contact the common event by placing the disk on two adjacent motifs, as illustrated in Fig. 4. For sliding a horizontal line-segment (the line-segment is not allowed to tilt), the test is to check whether the line-segment could contact the common event from above (See Fig. 5).

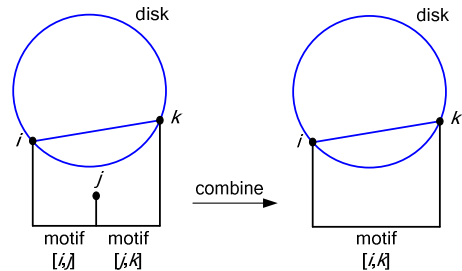


Figure 4. Motif combination by rolling a disk.

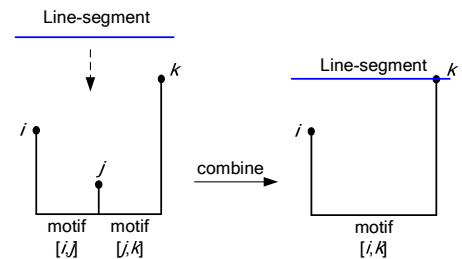


Figure 5. Motif combination by sliding a line-segment.

The motif combination procedure starts with the set of all events, namely all the sampled data on the profile,

and then it eliminates insignificant events by repeatedly applying the motif combination test on adjacent motifs until all the motifs could pass the test. The set of events on the motifs in the final solution are the discrete points that the disk could contact from above, namely the closing envelope. The pseudocode of the motif combination algorithm is presented in Fig. 6. The profile motif combination method results in a sequence of the points which may contact the structuring element when it is moving along the profile. With the final motifs, the closing envelope ordinates are computed by interpolating points on the arcs determined by the motifs at each sampling position for the circular structuring element, and finding the minimal height of two events on the motif for the line-segment structuring element, respectively.

```

Algorithm MotifCombination( $S, r$ )
{ Given a data set  $S$  with  $n$  points and the chosen }
{ disk radius  $r$ , compute the final motifs }

Chain list motifs =  $\{(p_1, p_2), (p_2, p_3), \dots, (p_{n-1}, p_n)\}$ ;
while 1
    if CombineMotifs(motifs,  $r$ )
        break;
    end if;
end while;
return motifs;

Procedure CombineMotifs(motifs,  $r$ )
flag = false;
motif1 = motifs(1);
for  $i = 2$  to motifs.length
    motif2 = motifs( $i$ );
    if CombineTest(motif1, motif2,  $r$ )
        motif1 = {motif1.Start, motif2.End};
        motifs.Remove(motif2);
        flag = true;
    end if;
end for;
return flag;

```

Figure 6. Motif combination algorithm.

#### IV. ALGORITHM DISCUSSION

The motif combination algorithm sets out to eliminate insignificant motifs and obtain significant motifs. It is an iterative method that the motifs are merged repeatedly until no more combination occurs. The final motif events are the contact points. This algorithm is coincident with morphological filters due to the fact that the features on the profile smaller than the structuring element in size are removed by the motif combination.

To evaluate the performance of the algorithm, it is necessary to analysis the time complexity of the algorithm. For the naive algorithm, the worse case is that the size of the structuring element is larger or equal to twice of the profile length. The calculation of each envelope ordinate involves the whole profile data, thus for worst case the time complexity is  $O(n^2)$ . For the motif

combination approach, the iterative process has the time complexity  $O(n)$ .

In order to verify the actual performance, experiments are carried out on the profile data with the point amount varying from 1000 points to 80,000 points. The profile data were sampled from a metal sheet surface. The evaluation length is 80 mm with the sample interval 1  $\mu$ m. The morphological closing filter with the 5 mm disk was performed on the profile data using the naive algorithm and the motif combination algorithm respectively. These algorithms were implemented by Visual Studio C++ and ran on a computer with 3.16GHz Intel Core Duo CPU and 3GB RAM. The performance data are listed in Table 1. It is obvious that the motif combination algorithm achieves much better performance than the naive method, especially in the case of large amount of sample points.

TABLE I. ALGORITHM RUNNING TIMES ON VARIOUS AMOUNT OF PROFILE DATA

Algorithm	5,000	10,000	40,000	80,000
Naive	0.0010s	1.0294s	4.8391s	9.9274s
Motif Combination	0.0076s	0.0157s	0.0609s	0.1238s

End distortion is common for the filtration of open profiles. Morphological filters are of no exception. For mean-line filters, a common solution is to add sufficient zeros to two ends of the profile, referred as zero-padding. Zero-padding is not suitable for morphological filters because the padded part of the profile should not be geometrically viewed as a horizontal line with zero height. Instead it should reflect geometrical features of the profile ends. To achieve this, the padding of the profile data is conducted by reflecting the two ends of the profile in the range of half size of the structuring element. Fig. 7 demonstrates the reflective padding of the profile data.

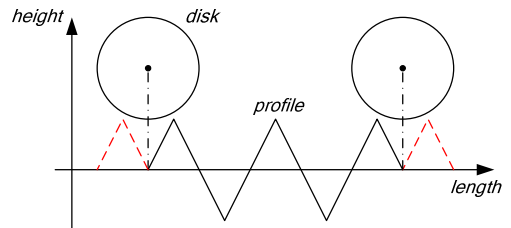


Figure 7. Reflective padding.

#### V. EXPERIMENTAL STUDIES

Fig. 8 and Fig. 9 illustrate two examples of applying the closing filter to a milled surface profile with the disk and the line-segment structuring element respectively. The experimental profile consists of 250 sample data, with length 1.25 mm and sampling interval 5  $\mu$ m. In Fig. 8, the profile is filtered by a 0.5 mm disk. The figure presents the closing envelope on the top of the profile, along with the solution motif events marked by dots. Fig. 9 presents the results of the morphological closing filter with line-segment length 0.1 mm.

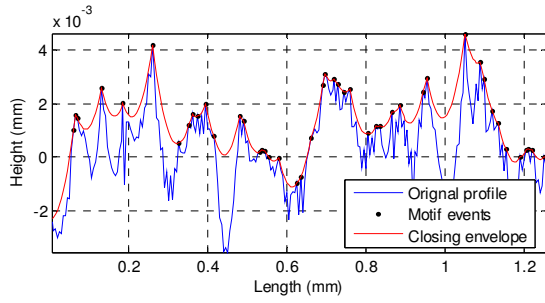


Figure 8. Closing envelope with disk radius 0.5 mm and motif events.

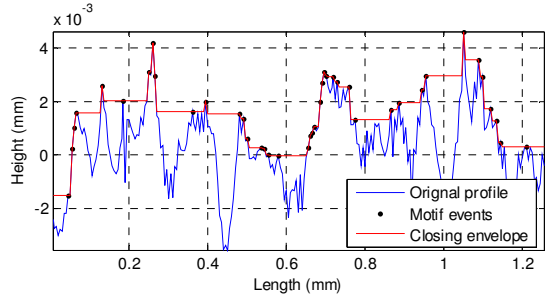


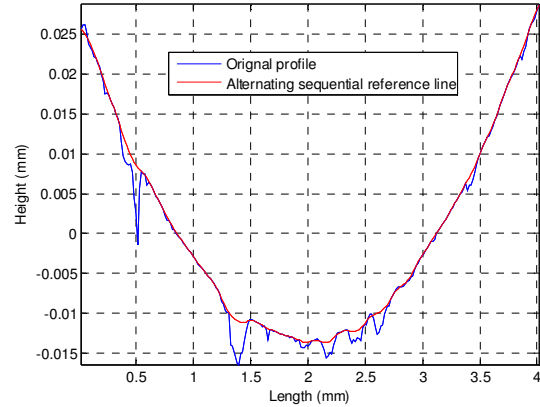
Figure 9. Closing envelope with line-segment length 0.1 mm and motif events.

## VI. EVOLUTION OF STRATIFIED FUNCTIONAL SURFACES

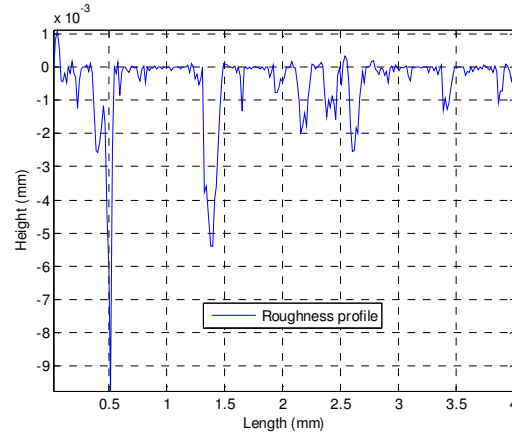
In engineering, surfaces with stratified functional properties are very common, for instance, the inner surface of cylinder liners. These kinds of surfaces are composed of deep valleys superimposed by plateaux. The plateaux support force bearing and friction while the valleys serve as lubricant reservoirs and distribution circuits. The traditional method for the analysis of these surfaces is performed by applying the two-stage Gaussian filter, the so-called Rk filter. However there are several drawbacks of this method [12]. Firstly, it was derived from empirical foundation with a significant assumption: surface contains a relative small amount of waviness, which is ambiguous and confusing. Secondly, running-in and running-out sections are generated from the Gaussian filter. These sections truncate the profile and only 20%-60% of the measurement data are used in evaluation.

In contrast, by using morphological filters, the profile does not need to be pre-processed to remove the form. The roughness profile can be obtained over the complete measurement length, therefore the resulting roughness profile has no running-in and running-out sections being “removed”. Fig. 10 presents such an example. The experimental profile was extracted from a plateau honed surface. It is obvious that the profile contains certain form component. Morphological alternating sequential filter, combination of first the closing filter and then the opening filter, with disk radius 5 mm, is employed to generate the reference line (See Fig. 10(a)). The closing filter suppresses all the valleys on the original profile that are smaller than the disk in size and the opening filter removes all the peaks on the resulting closing envelope which are smaller than the

disk. The roughness profile is obtained by subtracting the reference line from the original profile (See Fig. 10(b)).



(a) Original profile and reference line.



(b) Roughness profile

Figure 10. Roughness profile from an alternating sequential filter with disk radius 5 mm. (a) Reference line; (b) Roughness profile.

## VII. CONCLUSION

Regarded as the complement of mean-lined based filters, morphological filters provide better results for functional evaluation of surfaces. A practical algorithm is proposed to implement the morphological filters for profile data. The algorithm is based on the motif combination, which is an iterative process. The experimental data shows that it is superior to the naive algorithm in performance, with the time complexity  $O(n)$ . The end effect of morphological filters is corrected by the reflective padding. Both circular and horizontal structuring elements are available using this method. Two examples of applying the morphological closing filter with the disk and the line-segment structuring element on a milled surface profile are illustrated. Finally, the morphological alternating sequential filter is employed to evaluate the roughness of stratified functional surfaces. It has many merits over the tradition two-stage Gaussian filter. The profile does not need to be pre-processed to remove the form and the roughness profile can be obtained over the complete measurement length without truncated sections.

#### ACKNOWLEDGMENT

S Lou gratefully acknowledges to the University of Huddersfield for PhD Scholarship.

#### REFERENCES

- [1] X. Jiang, P. J. Scott, D. J. Whitehouse, and L. Blunt, "Paradigm shifts in surface metrology, Part I. Historical philosophy," *Proc. R. Soc.*, vol. A463, pp. 2071-2099, 2007.
- [2] H. Von Weingraber, "Zur Definition der Oberflächenrauheit," *Werkstattstechnik Masch Bass*, vol. 46, 1956.
- [3] V. Srinivasan, "Discrete morphological filters for metrology," *Proceedings 6th ISMQC Symposium on Metrology for Quality Control in Production*, TU Wien, Austria, 1998.
- [4] P. J. Scott, Scale-space techniques. *Proceedings of the X International Colloquium on Surfaces*, pp. 153-161, 2000.
- [5] C. M. Malburg, "Surface Profile Analysis for Conformable Interfaces," *Transactions of ASME*, vol. 125, pp. 624-627, 2003.
- [6] E. Decenciere, and D. Jeulin, "Morphological decomposition of the surface topography of an internal combustion engine cylinder to characterize wear", *Wear*, vol. 249, pp.482-488, 2001.
- [7] ISO 16610-49, Geometrical Product Specification (GPS)-Filtration, Part 49: Scale space techniques, 2010.
- [8] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1982.
- [9] M. Krystek, "Morphological filters in surface texture analysis", *XIth international colloquium on surfaces Chemnitz, Germany*, pp. 43-55, 2004.
- [10] ISO 16610-41, "Geometrical Product Specification (GPS)-Filtration, Part 41: Morphological profile filters Disk and horizontal line-segment filters," 2010.
- [11] P. J. Scott, "The mathematics of motif combination and their use for functional simulation," *Int. J. Mach. Tools Manufact*, vol. 32, pp. 69-73, 1992.
- [12] X. Jiang, "Robust solution for the evaluation of stratified functional surface," *CIRP Annals-Manufacturing Technology*, vol. 59, pp. 573-576, 2010.



# *Data Mining for Gearbox Condition Monitoring*

M. Baqqar, M.Ahmed , F. Gu

School of Computing and Engineering

The University of Huddersfield

Huddersfield, HD1 3DH, UK

Corresponding author: M.Baqqar@hud.ac.uk

**Abstract**— Engineering datasets have growing rapidly in size and diversity as data acquisition technology has developed in recent years. However, the full use of the datasets for maximizing machine operation and design has not been investigated systematically because of the complexity of the datasets and huge amounts of data. This also means that data analysis based on traditional statistic based methods are no longer efficient in obtaining useful knowledge from these datasets.

Thus this paper discusses dynamic and static datasets collected from a gearbox test rig with a typical drive system such that the datasets are considered representative for condition monitoring purposes. Dynamic datasets were analyzed to diagnose the condition of the gear: Healthy or Fault, using conventional signal processing techniques such as time-domain and frequency-domain analysis. The static data was also analyzed for comparative evaluation of detection performances.

This procedure of data collection and analysis allowed a full understanding to be gained of condition monitoring datasets and paved the way for developing a more effective Data mining approach and efficient database.

Moreover, to evaluate the effectiveness of using these new techniques, a prototype database was developed based on a gearbox test system and tested using these methods. The results obtained from a number of conventional methods have shown that data mining can obtain information for condition monitoring efficiently but not so accurately to give fault severity information, which is often sufficient for making maintenance decisions.

**Keywords**- *Gearbox condition monitoring, Data mining methods, Conventional methods* .

## I. INTRODUCTION

Condition monitoring (CM) is a technique for acquiring operating data and analyzing it to assess the health and condition of equipment. Thus potential problems can be detected and diagnosed at an early stage in their development, providing the opportunity to take suitable recovery measures before they become so severe as to cause machine breakdown. To obtain accurate results CM collects large amounts of data with wide diversity including operating parameters, high density dynamic signals and special event datasets to produce historical trends which are presented to engineers and stored in databases. This gives rise to the problem that the volume of data is very large and the relationship between measurements is very complicated. Consequently, the CM data is not always understood properly [1] and the extraction of useful and meaningful information from the data is extremely challenging. Because machines and sensor equipment are growing in complexity, combined

with the recent progress in information technology (IT), data acquisition systems (DQS) can produce an overwhelming amount of data which is continuously increasing and contains features representing hundreds of attributes. Data mining (DM) techniques based on Computational Intelligence, Machine Learning (ML) and advanced statistics have created new techniques and tools for automated extraction of implicit, previously unknown and potentially useful patterns and knowledge.

DM is an Artificial Intelligence (AI) powered tool that can discover useful information within a database that can then be used as a guide to action. The use of DM techniques in industry began in the 1990s [2-4] and has steadily attracted more attention so that now DM is used in many different areas in manufacturing to extract useful information for use in predictive maintenance, fault detection, quality assurance, design, production, scheduling, and decision support systems.

## II. GEARBOX CONDITION MONITORING

CM of a gearbox is a very important activity because of the importance of gears in power transmission in all manufacturing. Thus there has always been a constant pressure to improve the measuring techniques and analytical tools for early detection of faults in gearboxes. Gears are the most important element in the gearbox and, due to the high demands on them even at normal operating speeds and applied loads, gears are often subject to premature failure due to wear and material fatigue. To monitor the condition of a gearbox, physical parameters such as vibration, sound, temperature, even the motor current can be used.

Traditional methods of monitoring gearboxes are based on the assumption that any change in the condition of the gearbox may be detected by changes in the measured vibration signal [5]. This is because defects on a gear will alter both the amplitude and phase modulations of the gear's vibrations. Thus, any changes in vibration signal that can be measured and analyzed to provide an indication of the gearbox's condition.

The measured vibration signals often exhibit highly non-stationary properties, because the gear defects and incipient failures often show themselves in the form of changes in the spectrum of the signal. Thus, selection of the appropriate signal processing technique for detecting gearbox deterioration when the box is subjected to varying loads turns out to be crucial, since these vibration signals can also be heavily corrupted with noise, and are often non-stationary. The statistical parameters of the signal produced by the damaged component may not be altered

much by the presence of a transient defect especially in the presence of noise while varying loads may give a signal in the form of unexpected or uncertain sources. Thus false alarms will be generated and unnecessary maintenance cost can be incurred [6].

### III. CONDITION MONITORING USING DATASETS FROM A GEARBOX RIG

#### A. Data Characteristics

Gearbox data can be divided into two types: static datasets and dynamic datasets. As shown in the table, the static dataset contains mainly measurements from the controller and these were used to demonstrate the performance characteristics of the system. This type of data can give a quick indication of system health. The dynamic datasets are often used for CM as they can produce more details of the health condition and hence allow diagnosis of faults.

TABLE I. GEARBOX DATASETS

Static datasets	Dynamic datasets
Armature Current	Shaft speed
Load Set	Angular Speed
Speed Feedback	Motor current
Torque Feedback	Vibration signal from gearbox
Motor Current	Vibration signal from motor flange
Speed Demand	

#### B. Condition Monitoring Data Size

The amount of data acquired and stored for CM will be determined during the process of data collection via the data acquisition system. The memory size of the data collectors and data processors can affect the size of the collected data file. Data size and data points are very important factors in providing a clear picture of data collection. Table II shows size of the datasets collected from the gearbox test rig.

TABLE II. THE SIZE OF THE DATASETS COLLECTED FROM THE GEARBOX TEST RIG

Type of data	Size of data file
Static data	364 KB (373,079 bytes)
Dynamic data	152 MB (160,016,680 bytes)

#### C. Conventional Methods for Monitoring Gearbox

To better understand gearbox CM using traditional diagnostic techniques, experimental work was conducted on the same gearbox test rig as had originally been designed and previously used to monitor the health of a

two-stage helical gearbox using conventional techniques of vibration monitoring. The two-stage helical gearbox was used for the experimental work because, in addition to its widespread use in industry, such gearboxes allow faults to be easily simulated.

The goal of the experiment was to monitor the gear under healthy conditions and compare the results with those obtained when a realistic fault was introduced under similar operating conditions. Generally, many types of fault can be observed in gearbox operation, a broken tooth, a crack, scuffing, wear, etc. In this experiment, one broken tooth was seeded onto the 1st gear because it is a very common fault in gearboxes, and data was collected under different loads.

Conventional techniques used in this work include the waveform comparison in the time domain and spectrum analysis in the frequency domain. In addition, a direct comparison is made of the static data based CM.

### IV. STATIC MEASUREMENT BASED DETECTION

#### A. Detection from Static Datasets

From Figure 1, it can be seen that the load set was similar for both healthy and faulty conditions. There was a slight difference in the armature current but this difference is not sufficiently significant to be considered as a fault indicator.

Torque feedback gives the best indication of the presence of the fault, and it can be seen from the figure that there is an increase in the Torque feedback signals.

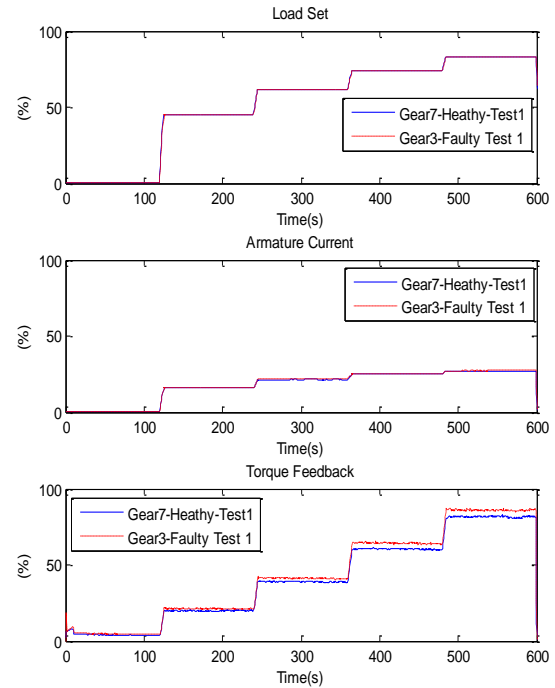


Figure 1. Load set, Armature current and Torque feedback for healthy and faulty gear conditions

From Figure 2 and 3, comparing the two conditions it can be seen that there is a slight difference between the healthy and faulty condition for all measurements.

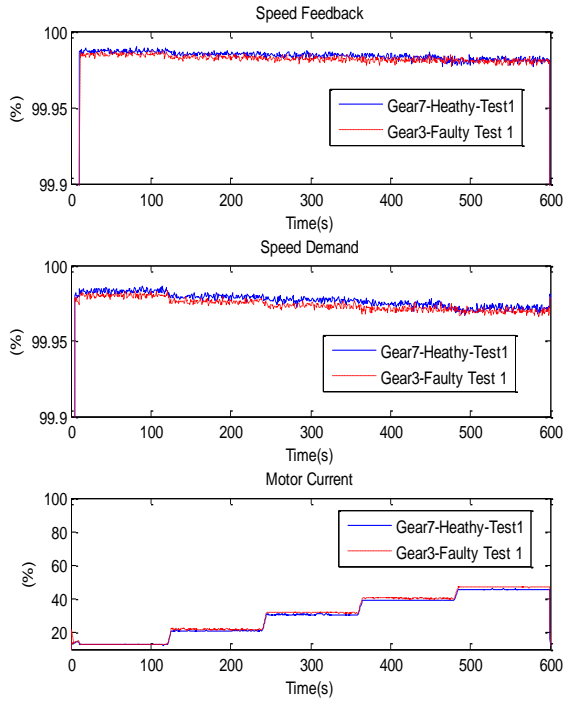


Figure 2. Speed Feedback, Speed Demand and Motor Current conditions for healthy and faulty gear

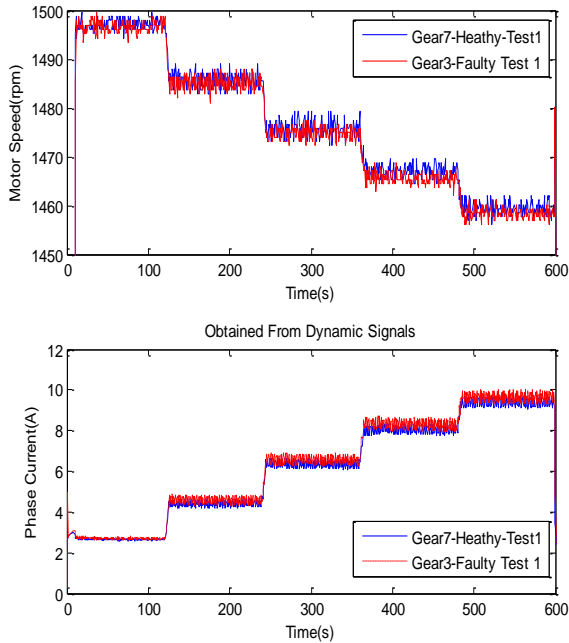


Figure 3. Motor Speed and Phase Current obtained from Dynamic Signal

## V. VIBRATION BASED DETECTION

### A. Detection in the Time-domain

Vibration waveforms for a healthy and faulty gear were collected from the accelerometer mounted on the gearbox casing with different current load operating conditions as shown in Figure 4. It can be seen that there are certain differences between the signal amplitudes due to the load variations for both gear conditions. Additionally, the waveform of the signals contains a massive amount of unknown information.

By comparing the two conditions it can be seen that all the waveforms of the signals exhibit some distortion but there are no clear fault indications even under high load conditions.

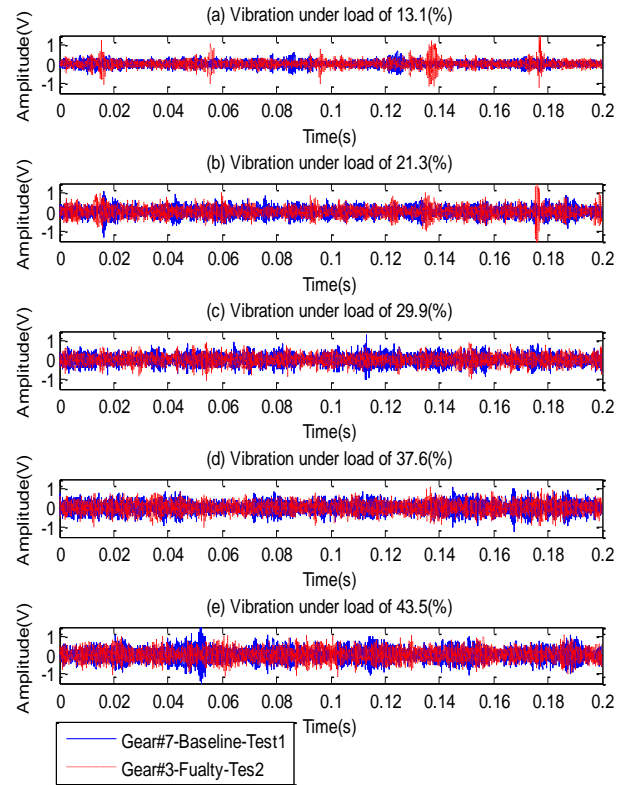


Figure 4. Vibration signals for healthy and faulty gears

For a more accurate clear and reliable assessment of the impact of load variation on the faulty gear, the RMS, Kurtosis and Peak values were found for the vibration signals.

From Figure 5, it can be seen that these statistical parameters varied with load, but there is no immediate or clear pattern to the fluctuations.

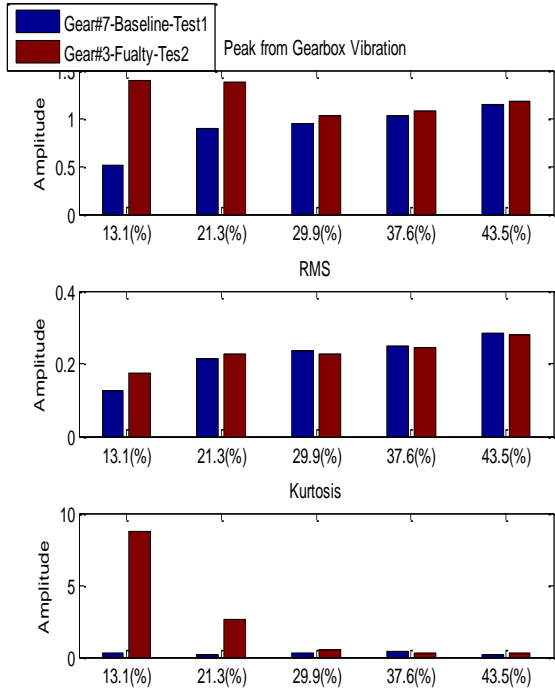


Figure 5. Peak, RMS and Kurtosis of time-domain vibration signal from gearbox

#### B. Detection in the Frequency domain

Frequency domain analysis for vibration signals of the healthy and faulty gear was carried out using the Fast Fourier Transform (FFT). The full spectrum of the vibration based on averaged healthy and faulty vibration signals are shown in Figure 6. Extracted from the figure are the dominant components in the spectrum are the transducer resonance responses.

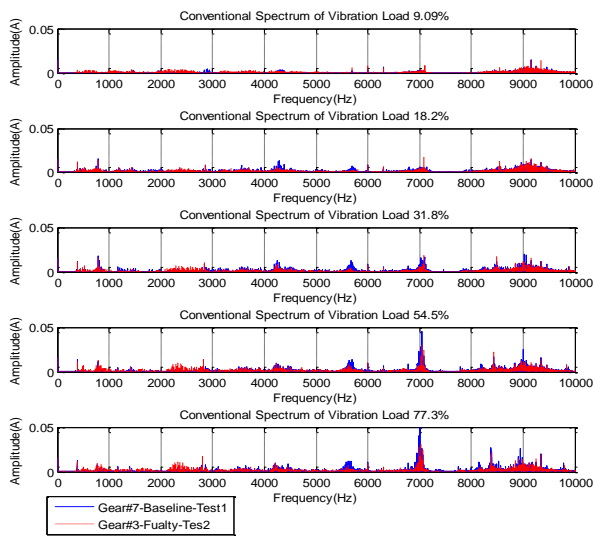


Figure 6. FFT spectral analysis of vibration signals from healthy and faulty gearbox

## VI. CURRENT BASED DETECTION

#### A. Detection in the Time-domain

Figure 7 shows distortion of the motor current signal – difference from a pure sine wave - but no clear symptoms of a fault appeared even under different loads.

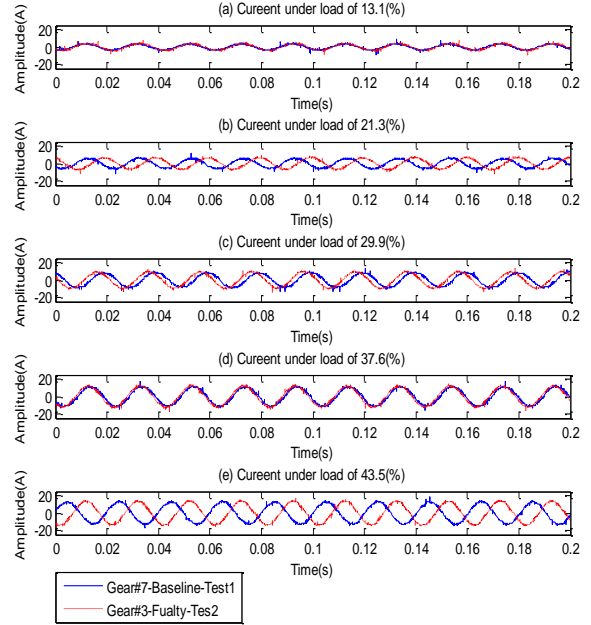


Figure 7. Stator current signals for healthy and faulty gears

From Figure 8, it can be seen that for all loads the Peak, RMS and Kurtosis for motor current are not significantly different for both healthy and faulty conditions.

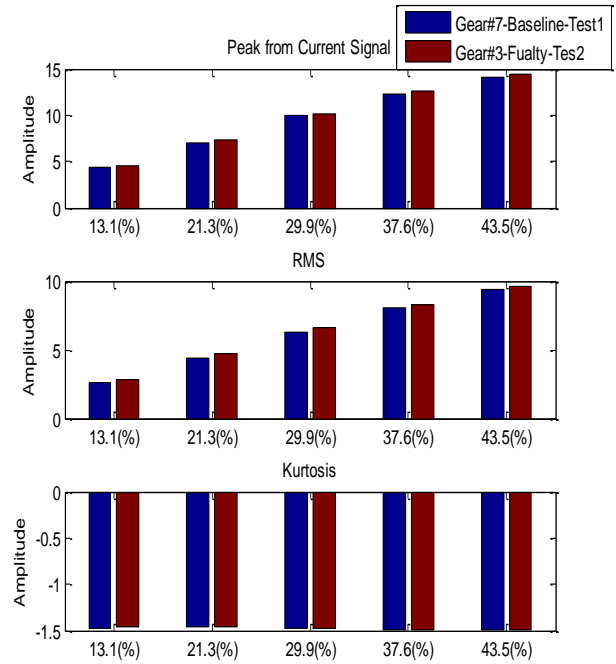


Figure 8. Peak, RMS and Kurtosis of stator current signals

### B. Detection in the Frequency-domain

It can be seen from Figure 9 that the vibration spectrum is rich with discrete frequency components (0 to 100Hz).

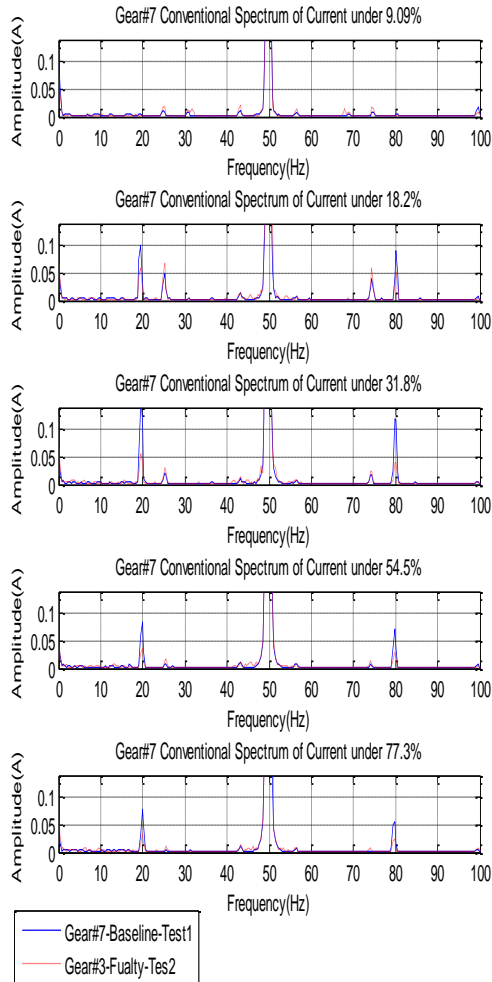


Figure 9. Spectral analysis of stator current signal measured for healthy and faulty conditions

### VII. CONCLUSIONS

The growing volume of CM data challenges conventional CM technologies and techniques in obtaining cost effective results. In parallel, DM technologies include neural networks, evolutionary algorithms, pattern recognition and support vector machines are also developing rapidly due to the advances in computing hardware. It is important to apply these technologies to the accumulated data to achieve more accurate and efficient CM.

The analysis of the dynamic datasets using conventional methods such as feature extraction from the time- and frequency-domains shows good detection and diagnosis results. However, the amplitude of the features is not sufficiently high for reliable diagnosis.

The primary study of the static data also shows certain detection information but not enough to identify the root cause of the faults.

### VIII. FUTURE WORK

- Pre-processing the dataset (choosing of data resources, cleaning the data from noise and errors, treatment unknown values, projection and reduction of data, etc.).
- Evaluating intelligent data mining techniques such as GA, NN, and SVM for Gearbox data.
- Selecting the DM task and technique for the extraction useful information and interesting and frequently occurring patterns.
- Build a mathematical model for vibration signal characterization under healthy and faulty gear condition; this requires the application of a suitable (mathematical and intelligent) data mining algorithms (NN), which describes the patterns.
- Validate the results of modelling based on the experimental data.
- Evaluation (test model, interpret and evaluate model).
- Post-processing the discovered or extracted patterns (further selection, ordering or elimination of patterns, visualization of the results).
- (Frequently some of the data mining DM steps need to be iterated several times to finally attain this goal).

### REFERENCES

- [1] S. McArthur; S. Strachan; and G. Jahn; "The design of a multi\_agent system for transformer condition monitoring," IEEE Transactions on Power System, vol. 19, no. 4, pp. 1845\_1852, 2004.
- [2] Lee, M. H , 1993, "Knowledge Based Factory," Artif. Intell. Eng, 8, pp. 109\_125.
- [3] Irani, K. B , Cheng, J, Fayyad, U. M., and Qian, Z , 1993, "Applying Machine Learning to Semiconductor Manufacturing," IEEE Expert, 8(1), pp. 41-47.
- [4] Piatetsky-Shapiro, G., 1999, "The Data Mining Industry Coming of Age" IEEE Intell. Syst, 14(6), pp. 32\_34.
- [5] Stander, C. J., Heyns, P. S. and Schoombie, W. 2002. Using Vibration Monitoring for Local Fault Detection on Gears Operating Under Fluctuating Load Conditions, Mechanical Systems and Signal Processing. Vol. 16, No. 6, pp. 1005-1024.
- [6] Zhan, Y., Makis, V. and Jardine, A.K.S., 2006. Adaptive State Detection Of Gearboxes Under Varying Load Conditions Based On Parametric Modelling, Mechanical Systems And Signal Processing, Vol. 20, No. 1, pp. 188-221.

# Process Monitoring and Metrology for Single Grit Grinding Test Performance

Tahsin Tecelli Öpöz and Xun Chen

Centre for Precision Technologies

University of Huddersfield

Huddersfield, UK

t.t.opoz@hud.ac.uk, x.chen@hud.ac.uk

**Abstract**— Single grit scratch test may provide better understanding of complex material removal mechanism of grinding process on the micro scale. In this paper, evaluation of single grit scratches was performed by utilizing monitoring and metrology devices. Particularly, AE and force sensors were used to monitor the process. AE sensitivity on material deformation was found comparable to force sensor sensitivity. High contact area interaction result in increase of AE raw signal amplitude. Grit cutting edge wear phenomena was also investigated under the digital microscope. Multiple scratches formation was observed in one rotation, due to grit cutting edge wear. AE sensitivity on scratch tests and grit wear are investigated in this paper to provide more insight into grinding process monitoring and material removal mechanism by single grit approach.

**Keywords** – Single grit test; grinding; process monitoring and metrology

## I. INTRODUCTION

Grinding is a material removal process to obtain desired product geometry with stringent tolerances together with high surface quality. It is particularly efficient in difficult to cut materials, such as ceramics or hardened steels. Grinding process is performed by utilizing a grinding wheel which is formed very large number randomly distributed abrasive grit with varying geometric orientation and size. Single abrasive grit performance on a work surface can be considered as an elementary action of grinding process. It is believed that the performing single grit test would be beneficial to assess the grinding operation in micro-level considering such as process mechanics and material removal mechanism, ground surface creation, etc. Material removal in grinding process can be accomplished by three postulated stages which are namely rubbing, ploughing and chip formation process. Rubbing has a negligible contribution into material removal since it only involves the elastic deformation in very small section of the grit-workpiece interaction area, whereas ploughing and chip formation both include plastic deformation of material and play main role in determination of ultimate surface formation and material removal from the workpiece. These three stages are difficult to study with grinding wheel, hence, single grit grinding test can be considered as a crucial test to understand the grinding performance in micro level. The early research on single grit grinding test was performed by Takenaka [1]. He observed cutting action in different depth of cut, even in

extremely small depth of cut lower than 0.4  $\mu\text{m}$ , and chips are observed in the form of tiny leaves torn from the workpiece surface. He also observed the three stages of material removal stages, namely, rubbing, ploughing and chip formation. Kannappan and Malkin [2] widely investigated the influence of grit size and operating parameters on the mechanics of grinding by using grinding wheel rather than single grit grinding. Ghosh et al [3] has studied the grinding mechanic by single grit test. They investigated the material ploughing and specific energy change with varying dept of cut. Doyle [4] investigated chip formation in single grit grinding test via quick-stop test to demonstrate material removal with abrasive particles having large negative rake angles. Feng and Cai [5] investigated the grinding velocity and grinding sectional area on the single grit grinding forces. They demonstrated the friction coefficient decreases with the increase of grinding speed.

To assess the process performance and to investigate the material removal mechanisms of grinding process in micro level, selection of the process monitoring and metrology tools as well as interpreting the acquired signals as clear as possible are essential. To obtain high confidence results, it is required to utilize high degree confidence and reliable sensory set for in situ monitoring. Acoustic Emission (AE) is one of the highly desirable monitoring tools in machining application since plastic deformation is one of the strong sources of AE. AE can be identified as transient elastic waves, from 25 kHz to several MHz, generated by the release of energy from localised sources of materials which are subjected to permanent alteration in their structure [6, 7]. Fracture is also possible source of AE since during the propagation micro cracks release elastic energy due to generation of new surfaces. Friction or rubbing between two surfaces is another potential source; surface asperities come into contact and are plastically deformed hence generating AE [7]. The superiority of AE over other sensor such as force and vibration is the acquiring high Signal/Noise ratio (S/N) at precision scale. Figure 1 shows the S/N ratio variation against uncut chip thickness in machining [8]. In this research, an AE sensor and a 3-axis force sensor are involved to monitor the single grit grinding process to get valuable information about process and to get more insight into grinding mechanic and process monitoring for further investigation. Besides, advanced metrology devices such as Talysurf CCI white light interferometer, Talysurf PGI stylus and Keyence Digital Microscope are



utilized to assess the form of grooves created by single abrasive grit.

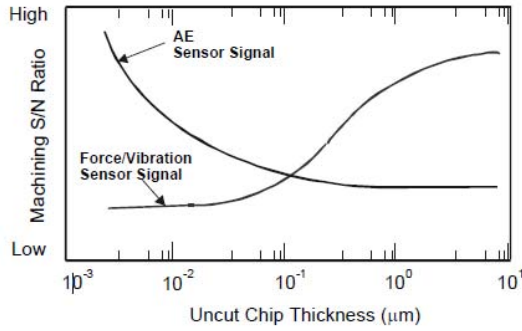


Figure 1. Signal/Noise characteristic of AE vs. Force/Vibration Sensors at Different Uncut Chip Thickness [8].

## II. SINGLE GRIT GRINDING EXPERIMENT

### A. Experimental setup

The experiments have been performed on a Precitech Nanoform250 ultragrand machine centre. A test rig was designed to facilitate the single grit grinding tests. CBN grit with a mesh size of 40/50 was used throughout the tests. The grit was glued onto the steel wheel with a diameter of 34.8 mm. Before grit attached to wheel surface, peripheral surface of steel wheel was ground after mounting to the work spindle by using high speed grinding spindle (at speed ~ 20000 rpm) to guarantee the concentricity of wheel with respect to machine spindle. Workpiece material was inconel 718 with a dimension of 30x40x10 mm in width, length and thickness. A Kistler 3-Axis force sensor was fixed between workpiece and designed workpiece holder. A Physical Acoustic AE sensor was installed onto the test rig, which was placed close to the workpiece. A closed up view of the test setup is shown in Figure 2. The workpiece surface flatness on the single grit test area was measured by LVDT probe and it was measured height difference between two ends around 11  $\mu\text{m}$  in horizontal direction as illustrated in Figure 3, which will provide increasing depth of cut when the wheel feed across the workpiece.

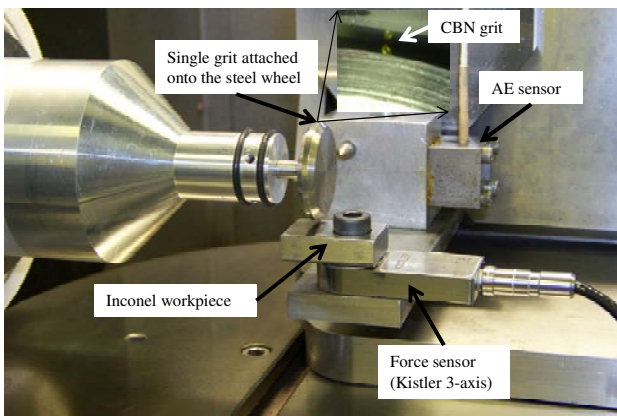


Figure 2. Experimental setup for single grit test on Nanoform250 ultra grind machine.

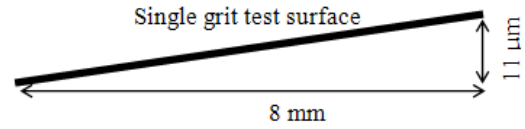


Figure 3. Single grit test surface height difference between two ends.

### B. Single grit experiment design

A series of single grit scratch experiment has been performed on inconel work material by using CBN grit. Before commencing the experiments, the CBN grit was ensured to touch around the middle of test surface by lifting workpiece manually using screw handle and then monitoring the touch by AE sensor. After touch, the grit was offset 2 mm to the left side of the workpiece (considering test setup in Fig. 2) to start the test. After the setup, the machine was allowed to run to perform single grit grinding by producing different scratches within an 8 mm width. This test repeated 10 times sequentially by shifting steel wheel position 2 mm in lateral direction. All tests' depth of cut was relatively different from each other because the CBN grit was manually set to touch the workpiece for each test. As such, this could not guarantee the same contact indentation depth. Tests conditions are given in Table I. Overall view from performed scratches are illustrated in Figure 4. Single grit scratches with a depth of cut of around 0.2  $\mu\text{m}$  and scratch length of 100  $\mu\text{m}$  can be obtained by using a slightly inclined workpiece surface together with fast table speed (Figure 5 and Figure 6).

Table I. SINGLE GRIT TEST CONDITIONS

	Wheel Speed (rpm)	Work Table Speed (mm/min)
Test-1	3000	200
Test-2	3000	400
Test-3	3000	500
Test-4	3000	500
Test-5	3000	500
Test-6	1000	300
Test-7	100	50
Test-8	2000	600
Test-9	500	200
Test-10	500	100

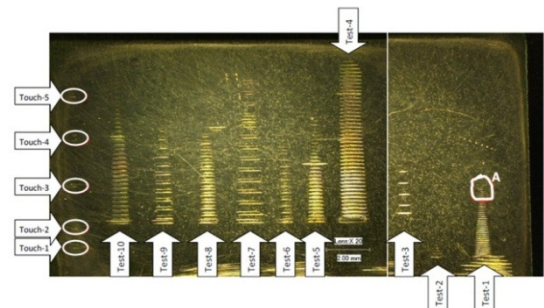


Figure 4. View of different scratches performed on Inconel 718 workpiece by using CBN single grit.

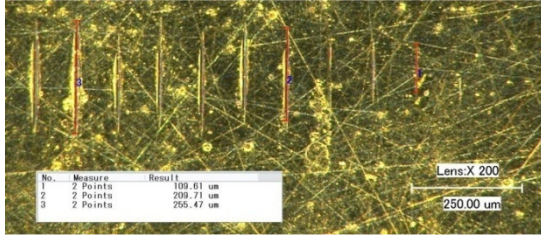


Figure 5. Magnified view of scratches in the area marked with 'A' in Figure 4.

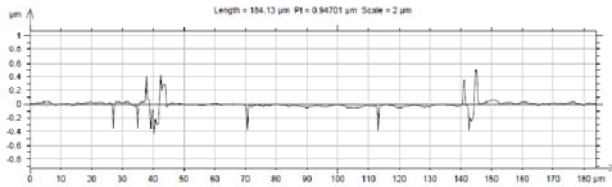
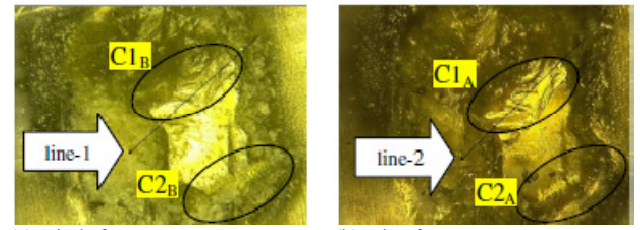


Figure 6. Scratches profiles (first two small scratches from right side of Fig. 5) measured by CCI interferometer.

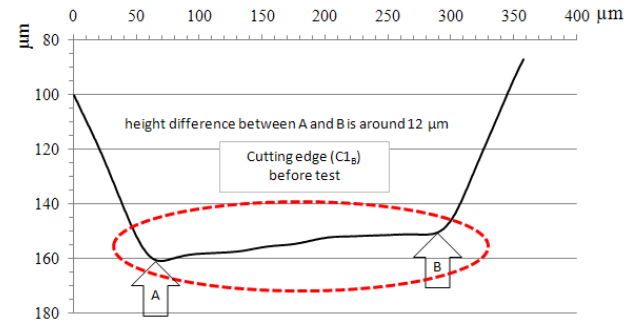
### C. Grit wear measurement

In grinding process, abrasive grit wear bear essential information which is strongly associated with the grinding performance. Attritious wear of the grit interacting surface with the workpiece turns to wear flat area along the grit cutting edges. Wear flat is one of the reasons of increase in specific energy of grinding process and may deteriorate the desired surface quality. In this research, 3D surface geometry of CBN grit is measured under Keyence Digital Microscope before (Fig. 7- (a)) and after (Fig. 7-(b)) tests. Grit region on the top of the grit represent the wear flat area, where the scratches were performed by these edges. 2D profile measurements show the two edge height difference  $|AB|$  is around 12 before tests (Fig. 7- (c)) and lower than 2  $\mu\text{m}$  after test (Fig. 7-(d)). The difference, 10  $\mu\text{m}$ , between these magnitudes represents the wear flat amount in 2D. Fig. 8 shows the two active cutting edges during the tests. In addition to cutting edge wear, single grit does not always generate single scratch, depending on the cutting edge distribution or cutting edge height variation throughout process, sometimes generate two or more scratches can be generated by single grit action as shown in Fig. 9 and Fig. 10. In the first 4 tests (test-1 - test-4), single scratch is generated by each rotation of wheel. In the middle of test-5, due to first cutting edge ( $C1$  in Fig. 8) wear, second cutting edge ( $C2$  in Fig. 8) come into contact with workpiece to generate secondary scratch in one rotation as shown in Fig. 8. At initial stage, secondary scratches are very small compare to primary scratch performed by first cutting edge, but at later tests, scratches performed by second cutting edge are growing with increasing contact area and indentation depth as shown in Fig. 10. The distance between scratches produced by two consecutive rotations is around 300  $\mu\text{m}$ . This is consisting with calculation depending on test conditions.

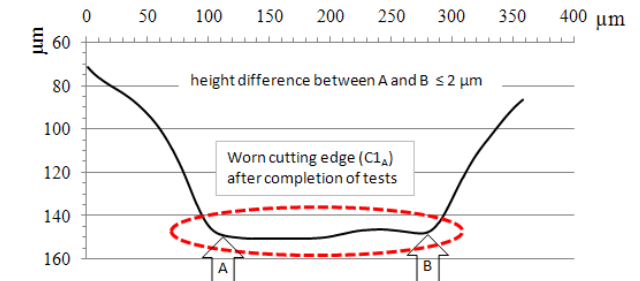


(a) grit, before test

(b) grit, after test



(c) Grit first contact (cutting) edge profile along line-1 drawn on (a), before tests performed.



(d) Grit first contact (cutting) edge profile along line-2 drawn on (b), after tests performed.

Figure 7. Grit cutting edges (a) before tests and (b) after tests test, wear flat measurement from profile (c) before tests and (d) after tests along the line drawn on (a) and (b) pictures.

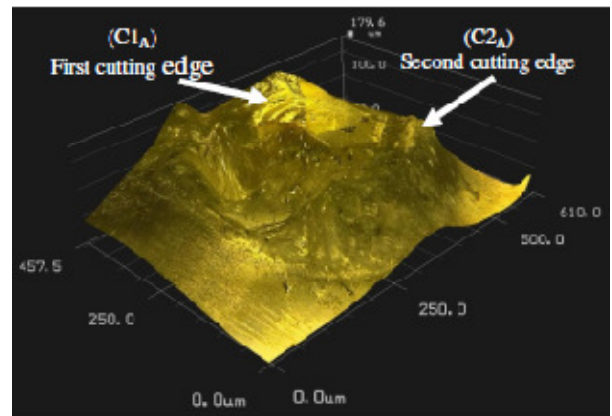


Figure 8. Grit 3D surface profile shows two separate cutting edges which were involved into scratch generation.



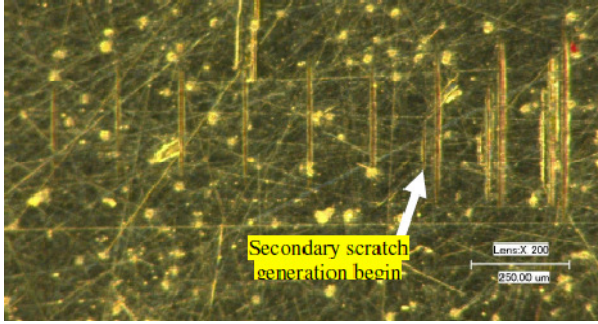


Figure 9 Secondary scratch generation start by a single rotation at test-5

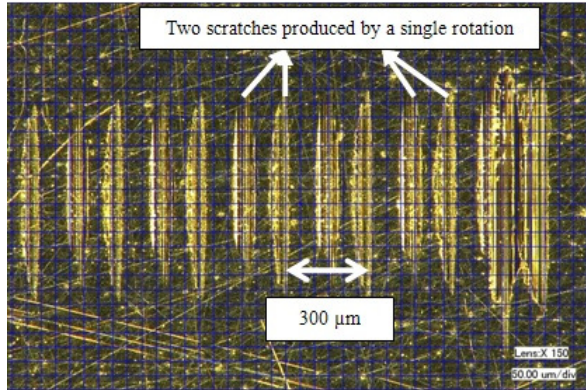


Figure 10. Double scratches by a single rotation from test-8.

#### D. AE and force monitoring and some discussions

It is believed that AE signal from machining source include valuable information about the machining process condition, material micro structure transformation, plastic deformation, fracture and so on. The critical thing is to interpret the AE signal properly utilizing some transformation technique. AE raw signal is stored in time domain and difficult to characterize it without transforming into frequency domain utilizing FFT or without RMS calculation [7]. In this paper, only raw signal is taken into consideration to demonstrate possible AE relation with the increase of contact area or depth of cut.

Scratches' depths of cuts are measured by CCI interferometer. Due to difficulty of measurement, some scratches from tests are selected to demonstrate possible relation of depth of cut with AE signal. AE raw signal amplitude does not show the linear relationship with the increasing depth of cut of scratches. This is possible owing to anisotropy of work material may leads to produce alternating elastic waves as it is the source of AE signals. Although AE signal magnitude increasing roughly when we consider increase in depth of cut in wide range but it is not sensitive for scratches having closer depth of cuts. If test-4 is handled to investigate, Fig. 11 shows the microscopic picture of scratches performed in test-4 and corresponding AE signal is depicted in Fig. 12. First scratch from left side in Fig. 11 has relatively smaller depth of cut and length. AE signal amplitude can be used to distinguish scratches' having high depth of cut difference. As it is evident from Fig. 12 AE signal increases abruptly with sharp increase in depth of cut, however AE signal fluctuate if the depth of cut increase slightly. Fig. 13 and Fig. 14 depict the power of

first and last AE signal in Fig. 12. These graphs show the first and last scratch has similar deformation characteristic although different depth of cut and different AE amplitude. Fig. 15-16 show the first and last part of scratches generated in test-10, respectively. Figure 12 demonstrate the corresponding AE signal monitored during test-10. AE raw signal fluctuate although increase when wide range in depth of cut taking into consideration. The power intensity of the first and last signal (Fig. 15-16) shows different characteristics, which are dissimilar to previous characteristic obtained from Fig. 13 and Fig. 14. The intensity of first signal (Fig. 18) very weak comparing to the last signals (Fig.19). When we consider the actions occurred during grinding process, the first scratch are formed by more rubbing and less ploughing action, whereas ploughing and cutting actions, involve heavily plastic deformation, are more prominent in the last scratch. To show the relational dependency of AE raw signal with depth of cut, some scratches depths are measured by CCI interferometer; it is difficult to measure the exact depth of cut so we can only measure approximate values.

In the present study, the forces generated by single grit actions are measured by 3-axis Kistler force sensor, but the evaluation of data is very obscure to define which force signal belongs to which scratch, since at some certain depth of cuts force sensor cannot detect force properly, signal dropped within noise level though it is filtered to diminish noise and increase readability of signal. Nevertheless, forces are obtained in higher depth of cut as shown in Figure 20 from test-4. The highest depth of cut occurred in the last scratch of test-4 is around 8  $\mu\text{m}$  and leads to approximately 3.8 N normal force and 1 N tangential force generation. However, when you come down from peak force to lower side (right to left in Figure 20 force is decreasing due to decreasing depth of cut but it is not easy to say exactly the smallest force caused by which scratch because the peak force is generated in the final scratch and final scratch involve several rotation without movement of wheel to stop the test. Furthermore, for the current study, oscilloscope is used as a data acquisition device, maybe sampling rate (833 Sa/sec when time division set to 1 sec) is not enough to pick up every signal generated during testing. For force analysis, some test should be re-performed to increase confidence and reliability.

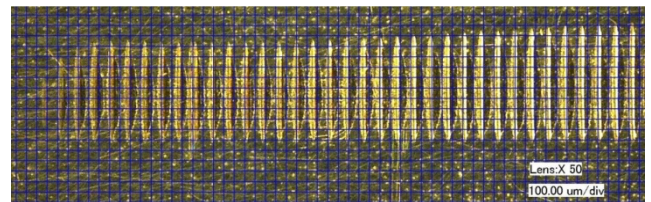


Figure 11. First 36 scratches (totally 45 scratches, last 9 not included due to space limitation) from test-4, the first scratch is the first one from the left side of the picture.

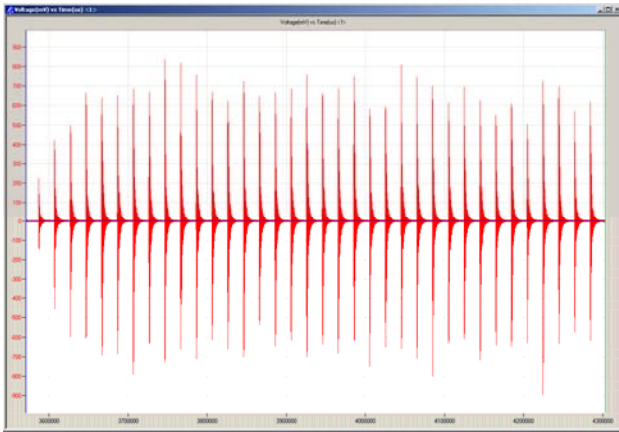


Figure 12. Raw AE signal (Voltage (mV) vs. Time (μs)) for scratches in Fig. 10.

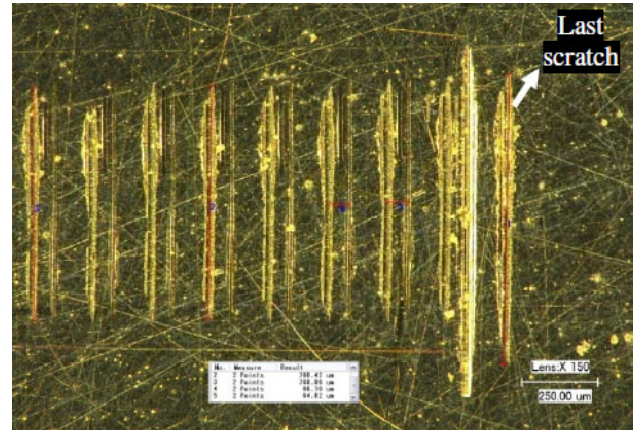


Figure 16. Last part of scratches performed in test-10.

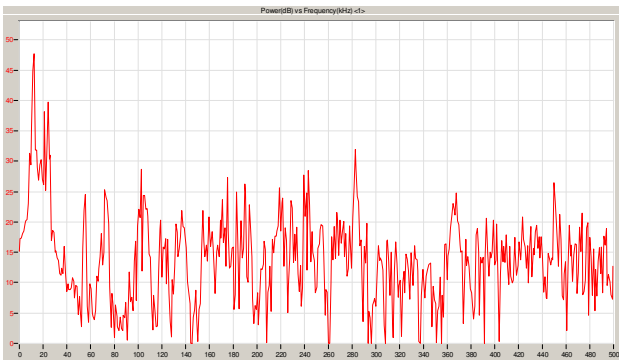


Figure 13. Power of first AE signal (dB vs. kHz) from Fig. 11.

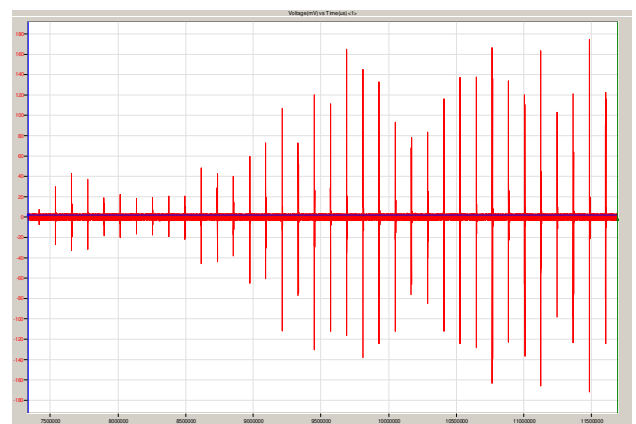


Figure 17. Raw AE signal (Voltage (mV) vs. Time (μs)) for scratches performed in test-10.

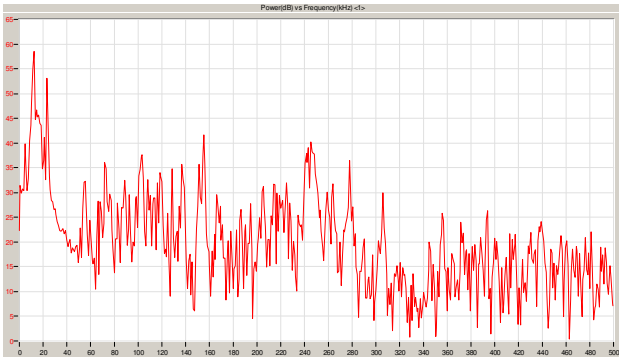


Figure 14. Power of last AE signal (dB vs. kHz) from Fig. 11

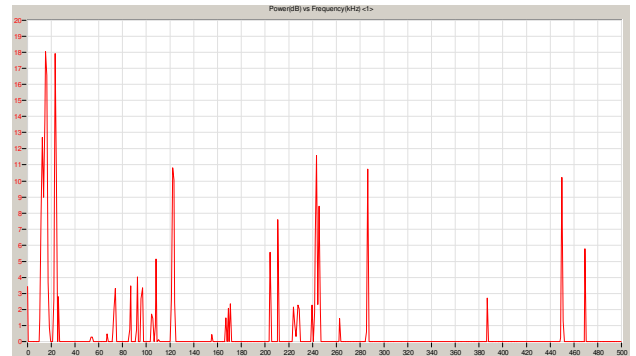


Figure 18. Power of first AE signal (dB vs. kHz) from Fig. 14.

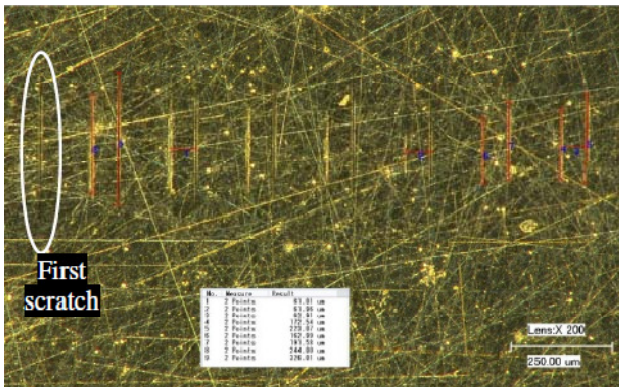


Figure 15. First part of scratches performed in test-10.

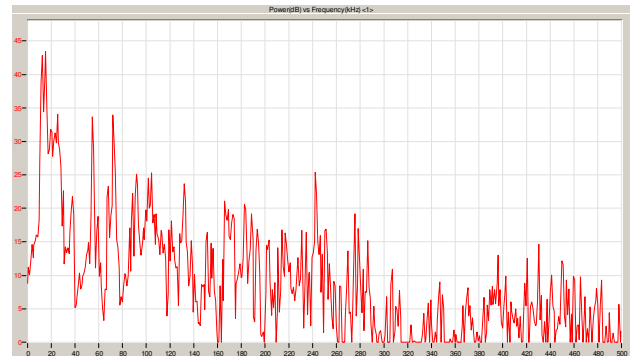


Figure 19. Power of last AE signal (dB vs. kHz) from Fig. 14.



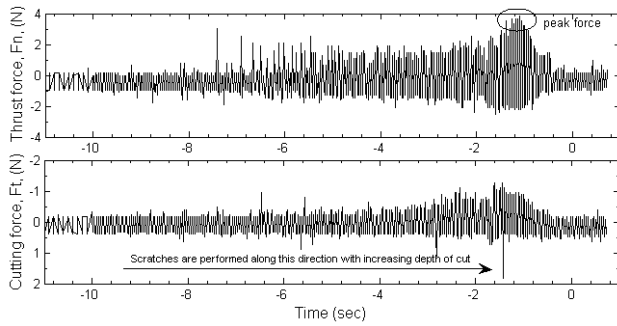


Figure 20. Force generation,  $F_n$  and  $F_t$ , during test-4.

### E. Scratch form evaluation

The scratches produced by single abrasive grit are composed of groove ground and material pile up across the side of grit interaction area due to severe plastic deformation occurrence. The rubbing stage of process is difficult to analyse due to only involving elastic deformation and it is recovered by elastic spring back effect. Chip formation would occur but this is also difficult to collect chips during testing. But eventual shape of groove and ploughed material around side of groove can be investigated by utilizing some advanced metrology devices. To evaluate scratch form, Talysurf CCI white light interferometer is used. After 3d non-contact measurement, 2d profile can be extracted from any direction and point. Fig. 20 shows profile extracted from 3d measurement of scratches from test-4. Scratches 1<sup>st</sup> and 2<sup>nd</sup> in Fig. 21 are the first two scratches performed in test-4, the form shape of scratches are quite clear, ploughed material and groove features can be distinguished clearly, but scratches of 21<sup>st</sup> and 22<sup>nd</sup> as shown in Fig. 21 are very close to each other due to increase of depth of cut, this leads to overlap of pile up material across the groove. 3d form measurements of scratches from test-1 are shown in Fig. 22 and extracted profiles from cross direction depicted in Fig. 23.

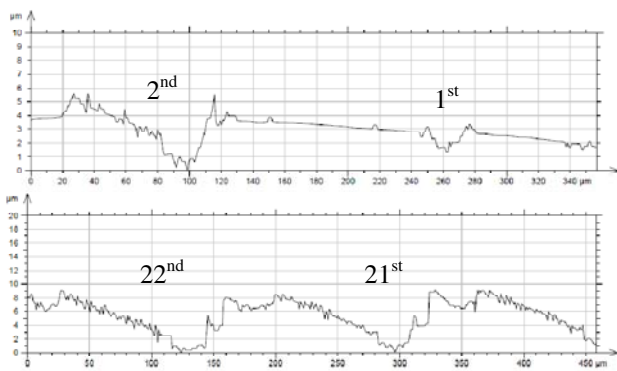


Figure 21. Scratches profiles by CCI measurement (numbers on figure show the scratches order) from test-4.

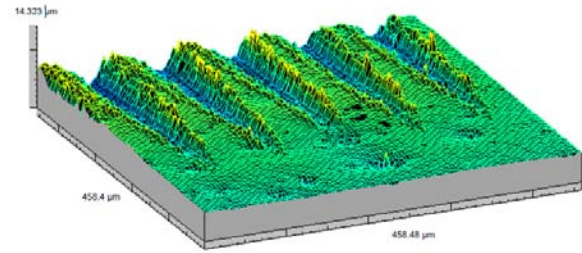


Figure 22. Half section of grooves ground by single grit, Talysurf CCI measurement for test-1 (scratches 6-10 numbering from left to right).

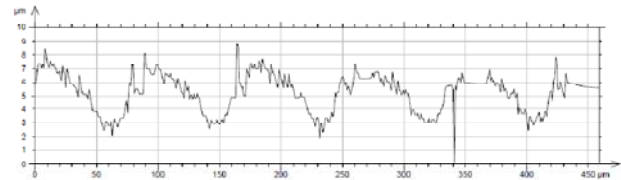


Figure 23 CCI profile extraction from Fig. 22.

### III. CONCLUSION AND FUTURE WORK

From this single grit grinding experiment, it can be observed that AE monitoring is an essential tool to characterize the process. The AE is sensitive enough to distinguish slight contact and heavy contact with the increasing depth of cut. AE raw signal amplitude increases with the increase of depth of cut. This may help to monitor grinding material removal process stages and distinguish them. Besides, single grit grinding experiment sometimes result in more than one scratch formation due to wear on cutting edge. Grit cutting edge and wear may also need further investigation to draw more credential consequences. Also, AE sensitivity against scratch speed needs further investigation to find more confident and reliable relation.

### REFERENCES

- [1] N. Takenaka, 'A study on the Grinding Action by Single Grit', Annals of the CIRP, vol. 13, 1966, pp. 183-190.
- [2] S. Kannappan, S. Malkin, "Effect of grain size and operating parameters on the mechanics of grinding," Journal of Engineering for Industry- Transaction of the ASME, vol. 94, 1972, pp. 833-842.
- [3] S. Ghosh, A.B. Chattopadhyay, S. Paul, 'Study of grinding mechanics by single grit grinding test', Int. J. Precision Engineering, vol. 1, 2010, pp. 356-367.
- [4] E. D. Doyle, 'On the formation of a quick-stop chip formation during single grit grinding', Wear, vol. 24, 1973, pp. 249-253.
- [5] B.F.Feng and G.Q. Cai, 'Experimental Study on the Single Grit Grinding Titanium Alloy TC4 and Superalloy GH4169', Key Engineering Materials, vol. 202-203, 2001, pp. 115-120.
- [6] A. Iturrospe, D. Dornfield, V. Atxa, J.M. Abete, 'Bicepstrum based blind identification of the acoustic emission (AE) signal in precision turning', Mechanical Systems and Signal Processing, vol. 19, 2005, pp. 447-466.
- [7] D.E. Lee, I. Hwang, C.M.O Valente, J.F.G. Oliveira, D.A. Dornfield, 'Precision manufacturing process monitoring with acoustic emission', Int. J. Mach Tool Manu, vol. 46, 2006, pp. 176-188.
- [8] D. A. Dornfield, 'Process Monitoring and Control for Precision Manufacturing', Production Engineering vol. 6, 1999, pp. 29-34.

# Improving Control Panel Consistency of Wizard-of-Oz Design and Evaluation Studies

Andol X. LI and John V. H. BONNER

Live:lab, Dept. Informatics  
School of Computing and Engineering, University of Huddersfield  
Huddersfield, UK  
[x.li@hud.ac.uk](mailto:x.li@hud.ac.uk) and [j.v.bonner@hud.ac.uk](mailto:j.v.bonner@hud.ac.uk)

**Abstract**—This paper investigates how a Wizard of Oz (WoZ) control panel could be developed to improve ‘between-subject’ consistency. To achieve this we conducted a comparative study of two control panels. Both control panels were used by the experimenter to ostensibly facilitate the design and evaluation of a novel domestic planning application allowing members of a family to coordinate a range of social arrangements and tasks. Based on video analysis and semi-formal interviews, the control panels as reliable design and evaluation tools were assessed. Results suggested that the component-separated control panel could obviously improve operational effectiveness thus enhancing system consistency.

*Wizard of oz, control panel, operation consistency, ambient intelligence*

## I. INTRODUCTION

This study is a part of our research into consistency problems using Wizard of Oz (WoZ) as a design and evaluation tool for potential future domestic communication applications. Since WoZ was coined by Kelley [1], it has grown in popularity as a user-based design and evaluation tool for unproven future technologies. WoZ can provide qualitative feedbacks of system interaction, acceptance and usability. However, it is difficult to provide consistent facilitations between subjects. Therefore, experimenter requires significant trainings to improve facilitation consistency. Some studies have attempted to address this problem by controlling system variables such as task complexities and participant preference [2]. Despite that, difficulties still remain in managing situated and unexpected decision making by study participants.

In this study our aim is to improve the control panel design for the WoZ system so that we can help experimenters to gather more consistent experimental data. To achieve this we conducted a comparative study which consisted of two different versions of control panel design. The control panels were used by experimenters to ostensibly facilitate the design and evaluation of the novel domestic planning application which allowed members of a family to coordinate a range of social arrangements and tasks. Through video analysis and semi-formal interviews the study assessed control panels’ effectiveness in terms of experimenter’s responding time, system error handling and participant engagement.

The paper is structured as follows: section 2 discusses the background to WoZ as a design and evaluation tool, considerations on consistency and control panel concerns with WoZ studies; section 3 discusses assessment criteria and study procedures; section 4 analyses the experimental data; section 5 covers the implications of results in terms of handling ‘system mistakes’ and response effectiveness, and finally conclusions and thoughts for future work are presented.

## II. BACKGROUND

### A. Wizard of Oz – why using this

Wizard of Oz (WoZ) is a light-weight HCI methodology which intercepts interaction between users and systems, and thus acts as an ‘intelligent’ device [3]. Experimenters facilitate input and output operations in order to deceive subjects into thinking the device possesses what levels of intelligence. Where in reality the intelligence is modelled and simulated by the experimenter rather than being programmed into the device [3]. This form of simulation provides low cost design and evaluation opportunities for smart products. It allows a wide range of interaction scenarios to be evaluated against acceptance and usability criteria without having to incur heavy development costs. WoZ method has been used for a multitude of applications, for example, the listening typewriter [4], speech-controlled telephone services [5, 6], natural body movement preferences [7], vision-based game controls [7], facial expression [8], handheld devices with animated characters [8, 9] and, environmental sensors for human interruptibility studies [10].

However one disadvantage of WoZ as a design and evaluation tool is that it is challengeable for experimenters (wizards) to present consistent responses. This problem can be found across a number of WoZ studies, for example, in simulating speech systems [2], multimodal systems [11] and intelligent agent systems [9]. These studies have put some considerations on addressing the problem by configuring system variables such as communication modalities [2] and participant dynamics [12]. Some other studies looked for solutions from cooperative interaction designs, such as, employing multiple experimenters to manipulate multimodal systems [13, 14]. In the Neimo project the study employed three experimenters to facilitate speeches, face and mouse recognition [11]. There are also some studies which



introduce automatic mechanisms for experimenters to facilitate system designs and evaluations, for example, training the wizard at well-defined tasks with ‘behaviour instructions [15]’ in [16].

### B. Addressing inconsistency problems

Solutions proposed by these studies are at high levels that some dynamic factors are concerned such as participant preferences. However, there is a pragmatic way to look for solutions from control panel designs which can be used for experimenters to facilitate the ostensible system. Due to control panel is the only tool for experimenters to facilitate the system, improving its design may effectively reduce operation inconsistency.

A few studies have looked into control panel designs for inconsistency problem. Decades ago the control panels were crude due to the system functions were simple [12]. To date with the growth of interactive media which makes multimodal user interfaces popular [11], WoZ is used for designs and evaluations of complex system. Salber and Coutaz (1993) employed multiple interfaces to address multimodal system facilitation. This solution introduced actually more experimenter dynamics into studies, and each experimenter required respective control panels with various complexities, which may deteriorate the inconsistency problem. Balbo, Coutaz and Salber (1993) suggested ‘predictive models’ for automatic evaluation systems to overcome the inconsistency problem which was caused by human experimenters. They proposed a WoZ platform to allow observations and automatic participant behaviour recordings. However their approach only reduced the inconsistency risks in data collection stage when interacting with multimodal interfaces for experimenter. Mavrikis and Gutierrez-Santos presented a ‘tapering’ approach to gradually replace a human experimenter as a computer for the design of intelligent learning environments [17]. In their iterative designs the control panel was transferred by computer programs in a gradual manner, the limitation was that only one facilitating function was considered each time though. In some speech-typing systems control panels were equipped with specific filters that inputs were intercepted by the control panel and then outputs were presented in machine-like manners [4, 5, 18]. And some other studies aimed at intuitive graphic interfaces for experimenters to manipulate the system such like gesture recognition systems [12], mixed reality systems [19] and human-robot interaction systems [15]. These WoZ studies mostly emphasised on control panel usability rather than on the modules and layouts to help experimenters improve the operation effectiveness and consistency.

In this regard, we proposed an empirical study to evaluate control panel designs which, probably, led to consistency improvement. We proposed two different control panel designs for experimenter to facilitate three independent system applications. By asking participants to use the novel system we gave experimenter different control panels to facilitate same experiment tasks. The comparisons between two experiments gave clues on what and why the control panel design could improve operation consistency.

### C. Control panels – how should these be designed?

In current WoZ studies end-user interface designs have gained more attentions than control panel designs for experimenter use such as speech telephone service designs [5, 18]. The control panel design should gain more considerations due to that it is the only tool by which experimenters can facilitate the system and present intelligence. In multimodal systems the control panel was split into sub panels for respective experimenters, and in single modal systems the control panel was first considered on its functions as Klemmer [18] and Whittaker’s [20] ‘wizard interface’ design. Some old-fashion control panel designs were simple, for example, Gould’s (1983) listening typewriter had only one textbox for speech transcription, and Höysniemi’s (1989) gesture recognition system had one simple interface for direction controls via keyboard. To date control panels are designed with multiple components such as buttons, textboxes and other graphical interface elements. These components have provided basic access to fast responses, although the delay that caused by experimenter is still noticeable as described in [7].

The control panel designs aim to associate the multifaceted and situated relations between system and experimenter [21]. Multiple roles of experimenter exist in WoZ studies, such as controller, moderator and supervisor [22]. These dynamics need to be addressed in control panel designs due to different experimenter roles require different control panels. For example, a supervision control panel has more surveillance functions yet with less control elements and a control panel has more manipulative components. Therefore in our study the control panels are designed on controlling purposes with multiple manipulative elements.

## III. METHODOLOGY

Our methodology was conceived as a practical, cyclical progress in which variables were controlled and compared throughout two experiments. Each experiment went through a set of tasks involving three system applications: a domestic calendar, a communication dialogue and a cube-based media manager. The domestic calendar was a speech-recognition application designed for family arrangements planning. The communication dialogue was designed for natural language dialogues and the media manager was a simple multimedia manager. A full experiment cycle contained three action stages - planning, acting and reflecting. The planning stage related to setting up experiments, the acting stage to conducting experiments procedures and the reflecting stage to data analysis.

All experiments were video recorded by a webcam which was set beside the experiment site. Data was manually transcribed into scripts for further analysis which was one significant part throughout our study. A semi-formal post-experiment interview was planned afterwards. Through interviews we aimed to extract user thoughts about system performances. In this WoZ study participant could not be told the truth until all experiments were completed. However such low-level ‘deception’ still

needed to be explained to participant for data use consents.

#### A. Control panel evaluation criteria

Criteria are important for control panel design assessments. Unlike normal usability evaluations, control panels in WoZ studies are invisible to participants. To evaluate these control panels we need to measure how the system performs while being controlled by experimenter. Therefore how participant reacts to the system can affect experimenter control panel operations.

There are two ways of reflecting experimenter operations, one is how ‘system mistakes’ are handled and the other is how real-time responses are presented. Fraser and Gilbert [2] suggested that the wizard should take account of making some mistakes to keep faithfulness to technologies, and they also suggested a 5 percent mistaken occasions. To fulfil this target the experimenter needs to make some errors on purpose. For instance, taking the speech ‘thunder’ as ‘honda’ [23] and then presenting a learning progress to convince participants that the system is automatically learning on its own. Some speech misunderstandings may also contribute to mistaken occasions since the experimenter cannot be able to comprehensively understand all speeches which are close to personal lives. By analysing participant reactions toward these errors, we can measure how flexible the experimenter-mediated system handles mistakes through control panels.

Responding in real-time is the other significant criteria aspect which reflects how fast the system can respond to participant. In this study these responses were all generated via the control panels, thus measuring how fast the system interacts with participants can reflect the extent of operation effectiveness. The response durations were gained from recorded videos according to time stamps.

#### B. Control panels – why united and split designs?

We designed the application system in the domain of domestic communication due to the home is an ideal site for emerging interactive technologies [24-26]. The three system applications were proposed based on daily routines in the home [27, 28], and these were integrated with mundane rhythms such like organising daily appointments and managing multimedia.

The system was programmed with C++ (MFC) in Visual Studio, therefore it had normal window elements. The system first used intranet-based communication for remote manipulation and surveillance. After several trial tests the system integrated control panels and applications in one computer due to the instable network could not afford simultaneous manipulation and video surveillance.

We proposed two versions of control panels. One was designed with united layouts which contain all modules in one panel (see Fig. 1), and the other design split the control panel into a group of sub-panels (see Fig. 2). This division was made according to system applications that each application had one separate control panel. Both control panels had the same application interfaces. The application interfaces were the same due to this makes

participant easy to compare system performance changes (see Fig. 3 and Fig. 4).

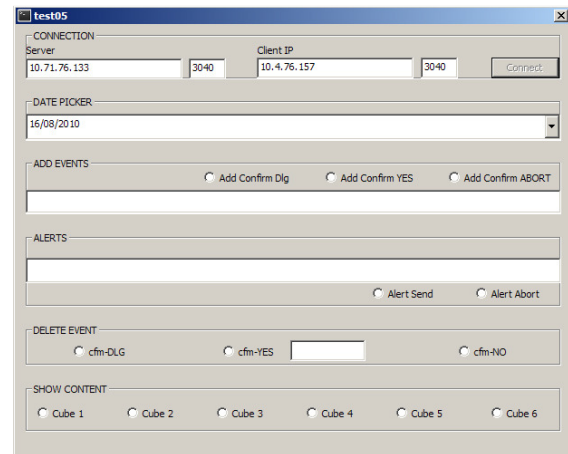


Figure 1. The compact control panel interface

#### C. Procedures – how experiments are conducted

Domestic interactions require experiment spaces which can provide home-like environments. We set up a scenario in laboratory as a domestic communication scene. It consisted of a set of sofa and a coffee table as well as multiple experimental devices such like projectors, webcams and microphones. The site separated some invisible space for experimenter to facilitate the system. Separated by big screens the experimenter could not be seen by participants, while the experimenter could still observe participant via surveillance video. The experimenter manipulated the host computer and monitored application running. The experimenter was located with the computer while application interfaces (Fig.3 and Fig. 4) were separately distributed in front of participant.

A volunteered participant was employed in our study. She had little knowledge of WoZ system but with strong interests in experiencing novel interaction styles. This was due to that experiments might be severely affected if participant became aware of experimenter existing. Furthermore the participant had good experiences of mundane affairs. With these conditions the participant was told that she was interacting with an intelligent computer system and her speeches could be recognised and learnt by the system. The participant needed to go through two experiments and she might be asked to compare the system performance changes in interviews.

The experimenter, who designed whole system applications as well as control panels, was playing the role of system facilitator. One advantage of this was that the experimenter did not require extra training to manipulate control panels. Meanwhile the experimenter could handle unexpected errors carefully based on system familiarity. The most important reason was that the experimenter could have first-hand experiences of control panel use.

The invisible experimenter could not directly associate with participant. To address that a new role was introduced which was called ‘instructor’. Her main responsibility was to deal with the participant face-to-face

as an experiment moderator. The instructor could deliver some indications about system applications.

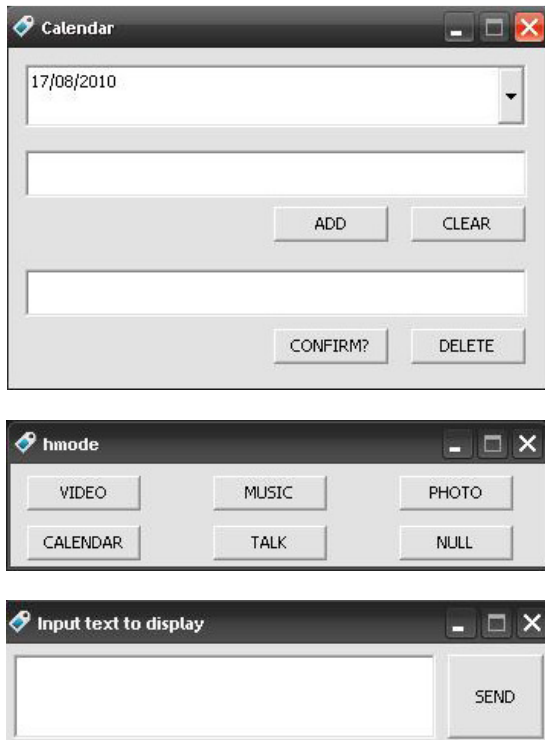


Figure 2. Split control panels: top - calendar, middle - media manager, and bottom - communication dialogue

Each experiment adopted a different control panel as described in Fig. 1 and Fig. 2. The first experiment started from system introduction by the instructor. After that the instructor gave participant a sheet in which command examples and tasks were listed. The instructor then allowed some time for participant to learn system functionalities. This was to make sure that the task completion was based on skilled manipulations, which might low the risks of dealing unfamiliar functionalities and wasting unnecessary time. Once the learning was done, participant was allowed to start tasks. All tasks should be accomplished and these included a) basic calendar operations (viewing / adding / deleting appointments), b) communicating with the system through the dialogue and c) using media manager to play videos. The media manager works with a coloured cube.

System responses were facilitated by the experimenter via control panels. When received incorrect speeches experimenter used the dialogue to show alerts. While using the cube the system launched a video and played it on the coffee table. The experimenter also needed to sense the cube movements which were assigned with different operations.

After all tasks the participant was invited to a semi-formal interview which encouraged the participant to express experiences and thoughts about the system. All comments made by participant were logged in videos. These videos were manually transcribed into scripts with time stamps, therefore operation durations could be calculated and analysed.

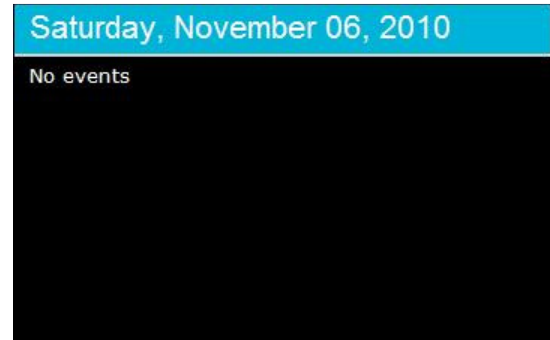


Figure 3. The domestic calendar application interface



Figure 4. The communication dialogue application interface

The participant was informed about the simulation system after the accomplishment of last experiment interview. And her consents were gained for further data analysis.

#### IV. EXPERIMENT ANALYSIS

The experimental materials were collected by observations and video analysis. The materials provided quantitative data in terms of response durations and mistake numbers, and also provided qualitative data which reflects how facilitation interacted with participant.

Videos were transcribed into texts by the experimenter who could recall the motivations of facilitation. These were complementary to understand participant reactions. Below is an example of scripts that demonstrates how the participant used calendar.

[00:05] Subject: today [*Speaking without hesitation and starting to wait for system responses*]

System: showing today's events

[00:08] Subject: create an event. [*Looking at the table and giving the speech when saw today's events*]

[00:10] System: popping up an input window for event contents

[*a short pause due to the participant was thinking about an appointment to add*]

[00:19] Subject: event for tomorrow and 2 o'clock

[*speaking naturally, still looking at the table and waiting for next response*]

[*A short pause due to the participant forgot to use the 'confirm' to finish the input*]

[00:22]System: popping up a message ‘Confirm?’

[00:23] Subject: ye, confirm [*suddenly realising the system would not respond without ‘confirm’ , then giving the right command and waiting for system responses*]

As two experiments followed same procedures the analysis compared some phenomenons in terms of real-time responses and ‘system mistake’ handling.

In the first united control panel experiment, the average response time was approximate 5 seconds, and the average speech typing duration was about 6 seconds. The speech typing was observed as the most time-consuming part in this experiment. The observation also showed that participant became impatient in some occasions for example system appeared out of response. When participant added a new appointment ‘Go Shopping’ the experimenter typed the speeches without responding to other inputs. In this case participant might repeat the speeches. However using the colour cube to manipulate videos did not have participant repeat speeches, and the communication dialogue neither.

In this experiment the experimenter facilitated some deliberate system mistakes. For example, in adding an appointment the speech ‘Go Shopping’ was mistaken as ‘Go For Shopping’; and in another case the system popped a prompt of ‘Speech not recognised, please repeat slowly’ when participant spoke, after a slower speech repetition the system again popped the reminding. Participant was observed to follow the system prompts.

In the second split control panel experiment the overall response time was reduced to about 3 seconds, and the average speech typing durations were still as high as 5 seconds. The split control panel was not observed to improve the typing speed while it noticeably reduced overall response time. Throughout the statistic we noticed that the massive time reduction of click operations contributed to the overall response improvement.

In this experiment same mistakes were avoided and some new mistakes were proposed. For example the system prompted ‘13<sup>th</sup> July’ instead of participant’s speech ‘30<sup>th</sup> July’. In this case participant repeated correct speeches until the system recognised them.

In both experiments there were response differences across system applications. It was observed that the colour cube manipulation had fastest responses and the domestic calendar had slowest responses. Meanwhile participant’s attitudes towards these applications were also different that communication dialogue gained most tolerances of slowness and the cube manipulation gained least. Through interviews the participant validated these observations.

## V. IMPLICATIONS AND DISCUSSIONS

Experiment comparisons explicitly illustrate the effectiveness differences between two control panels. Due to all experiment facilities were the same, the effectiveness improvement is deemed as a result of control panel design changes.

The reason of the split control panel having higher response effectiveness may be twofold, first the control panel can be rearranged according to participant preferences, and second the split may make control panel more intuitive. The rearrangement reduces unnecessary operations, for example moving cursor from left top to right bottom over the control panel, and also it allows participant to create an environment that fits better. The intuitive control panel may lower time waste on looking for functions. A complicated layout requires extra endeavours to locate a facilitating function. This may be even difficult when facing a control panel with massive boxes and buttons, such like Whittaker’s (2002) wizard interface designs.

However there is no strong evidence showing which control panel handles system mistakes better. The durations of generating and responding to system mistakes are similar in two experiments. One of reasons may be the same control panel elements through which experimenter can only facilitate limited mistakes. The only discovery is drawn upon system application differences in terms of user tolerance. Speech typing consumed most time while clickable elements (such like buttons and checkboxes) showed great potential to shorten response time in the second experiment. Therefore introducing clicking-style elements into control panel may help to handle system mistakes more consistently due to these elements provide fixed operations in fast response.

## VI. CONCLUSIONS

This paper presented a comparative study between two control panel designs to investigate how they should be developed for inconsistency problems in WoZ studies. The study built on control panel designs, and assessed operation inconsistency changes in terms of real-time responses and system mistake handling.

Based on video analysis and interviews, the study suggests that the control panel design can intuitively affect experimenter and thus affect system operation inconsistency. To address this issue, designs should consider the types of elements and manners of layouts. From this study it also shows that clicking-style elements are suitable for system mistake handling due to these provide limited yet fast responses, and separately grouped control panels are helpful for system responses as these help participant locate facilitating functions efficiently.

Despite this, we need further considerations on facilitating personal information-related intelligent systems in the future. This study merely considers control panels which receive participants’ data; future control panels should be designed capably to collect user information autonomously. Based on this, future studies will investigate some other principles for control panel designs with which experimenters can facilitate autonomous applications for domestic communication studies.

## REFERENCES

- [1] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," 1, ACM, 1984, pp. 26-41.
- [2] N. M. Fraser, and G. N. Gilbert, "Simulating speech systems," *Computer Speech and Language*, vol. 5, pp. 81-99, 1991.
- [3] A. X. Li, and J. Bonner, "Developing smart domestic applications using a wizard of oz methodology," in 2nd International Workshop on Human-Centric Interfaces for Ambient Intelligence, Nottingham, 2011.
- [4] J. D. Gould, J. Conti, and T. Hovanyecz, "Composing letters with a simulated listening typewriter," 4, ACM, 1983, pp. 295-308.
- [5] I. Bretan, A.-L. Eriback, C. MacDermid *et al.*, "Simulation-based dialogue design for speech-controlled telephone services," in Conference companion on Human factors in computing systems, Denver, Colorado, United States, 1995.
- [6] N. Dahlback, A. Jonsson, and L. Ahrenberg, "Wizard of Oz studies: why and how," in Proceedings of the 1st international conference on Intelligent user interfaces, Orlando, Florida, United States, 1993.
- [7] J. Hoysniemi, P. Hamalainen, and L. Turkki, "Wizard of Oz prototyping of computer vision based action games for children," in Proceedings of the 2004 conference on Interaction design and children: building a community, Maryland, 2004.
- [8] T. Bickmore, "Towards the design of multimodal interfaces for handheld conversational characters," in CHI '02 extended abstracts on Human factors in computing systems, Minneapolis, Minnesota, USA, 2002.
- [9] D. Maulsby, S. Greenberg, and R. Mander, "Prototyping an intelligent agent through Wizard of Oz," in Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems, Amsterdam, The Netherlands, 1993.
- [10] S. Hudson, J. Fogarty, C. Atkeson *et al.*, "Predicting human interruptibility with sensors: a Wizard of Oz feasibility study," in Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, Florida, USA, 2003.
- [11] D. Salber, and J. Coutaz, "Applying the Wizard of Oz technique to the study of multimodal systems," *Human-Computer Interaction*, pp. 219-230, Berlin/Heidelberg: Springer Berlin/Heidelberg, 1993.
- [12] A. G. Hauptmann, "Speech and gestures for graphic image manipulation," in Proceedings of the SIGCHI conference on Human factors in computing systems: Wings for the mind, 1989.
- [13] L. Molin, "Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces," in Proceedings of the third Nordic conference on Human-computer interaction, Tampere, Finland, 2004.
- [14] S. Balbo, J. Coutaz, and D. Salber, "Towards automatic evaluation of multimodal user interfaces," in Proceedings of the 1st international conference on Intelligent user interfaces, Orlando, Florida, United States, 1993.
- [15] Y. XU, S. TAKEDA, and T. NISHIDA, "A WOZ Environment for Studying Mutual Adaptive Behaviors in Gesture-based Human-robot Interaction." pp. 40-46. in AAAI 2007, Human Implications of Human-Robot Interaction Workshop, Vancouver, British Columbia, Canada, 2007.
- [16] Y. Xu, K. Ueda, T. Komatsu *et al.*, "WOZ experiments for understanding mutual adaptation," *AI & Society*, vol. 23, no. 2, pp. 201-212, 2009.
- [17] M. Mavrikis, and S. Gutierrez-Santos, "Not all wizards are from Oz: Iterative design of intelligent learning environments by communication capacity tapering," *Computers & Education*, vol. 54, no. 3, pp. 641-651, April 2010, 2010.
- [18] S. R. Klemmer, A. K. Sinha, J. Chen *et al.*, "Suede: a Wizard of Oz prototyping tool for speech user interfaces," in Proceedings of the 13th annual ACM symposium on User interface software and technology, San Diego, California, United States, 2000.
- [19] S. Dow, J. Lee, C. Oezbek *et al.*, "Wizard of oz interfaces for mixed reality applications," in CHI 2005, 2005.
- [20] S. Whittaker, M. Walker, and J. Moore, "Fish or fowl: A wizard of oz evaluation of dialogue strategies in the restaurant domain," in Language Resources and Evaluation Conference, 2002.
- [21] A. Voss, M. Hartswood, R. Procter *et al.*, "Introduction: configuring user-designer relations: Interdisciplinary perspectives," *Configuring user-designer relations*, A. Voss, M. Hartswood, R. Procter *et al.*, eds.: Springer, 2009.
- [22] S. Dow, B. MacIntyre, J. Lee *et al.*, "Wizard of Oz Support throughout an Iterative Design Process," 4, IEEE Educational Activities Department, 2005, pp. 18-26.
- [23] A. X. Li, and J. Bonner, "Enhancing social relationships through human-like intelligences." in The 9th International Workshop on Social intelligence Design (SID2010), Egham, UK, 2010.
- [24] A. F. Blackwell, J. A. Rode, and E. F. Toye, "How do we program the home? Gender, attention investment, and the psychology of programming at home," *International Journal of Human-Computer studies*, vol. 10, no. 1016, 2008.09.11, 2008.
- [25] J. Hughes, J. O'Brein, and T. Rodden, "Understanding Technology in Domestic Environments: Lessons for Cooperative Buildings " *Cooperative Buildings: Integrating Information, Organization, and Architecture*, pp. 248-261: Springer Berlin / Heidelberg, 2007.
- [26] R. Harper, *Inside the smart home*, New York: Springer-Verlag, 2003.
- [27] A. S. Taylor, R. Harper, L. Swan *et al.*, "Homes that make us smart," *Pers Ubiquit Comput*, 2007.
- [28] A. Crabtree, and T. Rodden, "Domestic routines and design for the home," *Computer Supported Cooperative Work*, vol. 13, pp. 191-220, 2004.

# Assessing customized product design using virtual human and imposed motion

Shengfeng Qin<sup>1</sup>, George Panayiotou<sup>1</sup>, Pin Zhang<sup>2</sup>

<sup>1</sup>School of Engineering and Design, Brunel University, UK

<sup>2</sup>School of Arts and Design, Nankai University, China

Sheng.feng.Qin@brunel.ac.uk

**Abstract**—This paper presents a new method of using digital human with imposed motion and motion dynamics to evaluate product design. We have demonstrated it with a case study. The various design concepts were physically mocked up as testing prototypes (scenarios). The users were required to mimic the use of the products and their interaction motions with the test prototypes were captured, recorded and analyzed. Then, the users were scanned with a 3D body scanner to help the creation of the corresponding virtual human models, which had the same (or very similar) body sizes to the real users. Finally, the virtual human models with corresponding motions were integrated in a computer aided ergonomics analysis system-UGSTM JACK<sup>®</sup> to review and evaluate different product design concepts. In addition, motion dynamics information was used to help conceptual selections. The proposed method combines the advantages of both real task motion and digital human simulation. With this method, the task motion is real instead of synthetic one. In ergonomic assessment software such as JACK<sup>®</sup>, there is no need for creating a CAD-based virtual product model to generate synthetic motions. This method can support ergonomic evaluations of product design especially haptic interface design at the conceptual stage without its CAD models.

**Keywords**- *Digital human modelling; ergonomic assessment; customised product design and development.*

## I. INTRODUCTION

Under the pressure of global manufacturing and customized product development, how to assess a conceptual design ergonomically is becoming an important issue since decisions made during this stage have significant influence on the cost, lead time, performance, reliability, safety and environmental impact of a product. Even the highest standard of design detail can never compensate for a poor concept. It is believed that ergonomic analysis using digital human modeling and simulation method can shorten the design time, lower development costs, and the design options can be assessed proactively [1].

Digital human modelling (DHM) includes gender, the skeleton system, body dimensions, appearance, vision, range of motions for each joint, biomechanics model, and so on. Several commercial software systems [2] such as JACK<sup>®</sup>, RAMSIS<sup>®</sup>, Safework<sup>®</sup> are available for creating virtual humans and ergonomic studies for product and

workplace design. Nevertheless, these ergonomic assessments are based on predicted task postures by engineers or ergonomists. Commercial products do not consider all the possible interactions with products. For example, JACK<sup>®</sup>, a computer-aided ergonomics tool, enables users to position bio-mechanically accurate digital humans of various sizes in a virtual environment, and to assign them tasks and analyze their performance. For some interactions, the software users need to manipulate a virtual human model into a specific posture, which is a time-consuming job. The effectiveness of ergonomic analysis with digital human models not only depends on the accuracy of a real user (group) being modeled, but also the interaction motions being assigned to the digital human. For a static analysis, a whole body motion is segmented into a series of specific body postures; then each posture can be used for a static ergonomic evaluation [3]. In customized product design application, there are two research problems: (1) how to create a best match between a virtual human to a real user (group) and (2) how to obtain a realistic motion to reflect the user's interaction to the designed product especially when haptic interfaces are needed.

In this study, we propose a new method to perform ergonomic assessments by a digital human with real task motion data which can reduce the deficiency of predicting task postures by engineer or ergonomists. Motion capturing and body scanning techniques are integrated to collect body data and motion data accurately. Based on accurate body measurements of a real user, a digital human is modeled in JACK<sup>®</sup> software. In order to obtain realistic task performance motions for the digital human, a physical mock-up (mainly the interface parts) of a product conceptual design is first built up and then a real user (a representative of a user group) is required to perform interaction tasks. The interaction process is captured by a motion capture system. The user's motion is then imposed to the digital human in JACK<sup>®</sup> to conduct various ergonomic assessments. In addition, the motion dynamic features such as the task performance time and motion changes are used to help select design concepts. With this method, we can create a best match between real human and digital human. The task motion is real instead of synthetic one. In ergonomic assessment software such as JACK<sup>®</sup>,



there is no need for creating a CAD-based virtual product model to generate synthetic motions because a digital human with imposed real motion is enough to perform an ergonomic assessment. Furthermore, this method is especially suitable for ergonomics purposes when haptic interfaces are needed. This method can support ergonomic evaluations of product interface design at the conceptual stage.

## II. RELATED WORK

Digital human modeling technology has been used for product design, workplace assessment, digital factory, medicine and entertainment [4]. There are two typical methods of using DHMs in ergonomics assessment. In the first method, virtual humans are created in a commercial software system and then placed in a virtual environment generated either inside the digital human system or imported from a CAD system [5]. Within the virtual environment, digital humans are assigned to perform a series of specific tasks with manipulated motions. For manipulation, the system designer can control DHM motion by forward and inverse kinematics. Actually, the manipulation is just a simulation of real human motion, but it lacks reality. A new digital human environment-Santos<sup>TM</sup>, has a dynamic motion prediction capability based on optimization techniques [2]. While Chaffin [6] argued that existing posture and motion prediction models now used in DHMs must be based on real motion data to assure validity for complex dynamic task simulations. Similarly, an online system has been demonstrated [7]. One of the problems in this method is that the designers have difficulties in predicting how a person interacts with the virtual workplace [8]. This in turn leads to difficulties in evaluating product design and ergonomic evaluation. It is the intent of the motion capture technology to capture the real working postures and motions to enhance the reliability of using DHM in ergonomic task evaluation and product design.

The second method of using DHM is virtual interactive design. It provides information about the human-machine or human-system interaction. First, virtual product and environment need to be built up, which can be perceived through virtual reality equipment such as VR head-mount design immersion. Then the real human interact with the virtual world (environment) via VR devices and their motions are captured with a motion capture system. Finally, the virtual environment and a digital human are created in a computer-aided ergonomics analysis tool to conduct design interaction review and ergonomics analysis. In this method, real human motion of interacting with virtual prototypes (environments) is used more realistically compared with the manipulated motion. This method has been successfully applied to an Automatic Teller Machine (ATM) design [9]. When considering VR-based perception, the user needs, for example, to wear head-mount immersion equipment. The user's motion of interacting with virtual objects may be different from that of real 3D objects. This method may be good for evaluating a product design when its visual appearance and perception is a major influence factor to user interaction.

When the ergonomic assessment involves haptic interface design, neither of the above methods is suitable no matter using virtual human and virtual objects or using real human and virtual objects because the haptic interaction is difficult to realize in those settings. Chang and Wang [10] showed an example of using a real product assembly line and captured assembly motion for workplace evaluation.

There are a big body of studies on motion capture techniques such as motion retargeting and motion synthesis for the creation of films and animations [11], which are not related to our research. In contrast, only a few studies focused on product design applications. Qin et al [12] used a physical reference prototype and hand motions for large surface design and Qin et al [13] proposed a design-by-motion method for collaborative design and surface modeling. Yi et al [14] combined hand gestures and hand motions into architectural design application.

The 3D body scanning technique has been used for apparel design and human engineering [15][16]. It has been rarely used for ergonomic analysis.

Comparing to the above, the proposed method uses physical design prototypes to capture real user task performance motion with a motion capture system. Several traditional ergonomic analysis methods embedded in JACK<sup>®</sup> are used to evaluate design interactions and product design. The method combines the advantages of both real task motion and digital human simulation.

## III. PROPOSED NEW METHOD

The new method aims to integrate a more accurate digital human model with a real user's performance postures (motions) into product design evaluation. It has two parts as illustrated in Figure 1. The first part involves user selection, 3D design prototyping and user testing. After selecting participants, their body information are collected with a [TC]<sup>2</sup> 3D body scanner in order to create better matched digital humans for ergonomic analysis later on. In parallel, different product design options are mocked up as 3D prototypes. Then, participants perform user tests with the 3D prototypes. While they perform their tests, their 3D body motions are captured with a 3D motion capture system for both visual design evaluation via 3D motion dynamics analysis and imposing motion to virtual humans for ergonomic analysis. All tasks above are conducted in real 3D space. The second part is to create virtual humans with better matched body sizes of real users (participants) and impose the captured motions to the virtual humans for ergonomic analysis. A virtual environment (optional) is created for virtual task simulation. It provides a virtual presentation of a design option.

In this method, a selected user is required to conduct a task, for example, picking up a kettle and putting it on a table. The prototype of the design (the kettle and the table) needs to be placed in the scene of a motion capture system (MotionAnalysis<sup>®</sup>). The user interaction to the product (the kettle) and its motion is captured through reflective markers attached to the participant's body. Because we

know the real user's task motions, we do not need to assign any posture and task to a virtual human in the JACK<sup>®</sup> environment, aiming to obtain a simulated task motion. Instead, we need to enable a peer-to-peer communication between the motion capture system and the JACK<sup>®</sup> system. In JACK<sup>®</sup>, a virtual human figure can be created matching with the real user (Fig 2). Also, in the motion capture system, a skeleton model of the real user can be generated from a given marker set and motion. When two systems communicate with each other, the virtual human in JACK<sup>®</sup> is controlled and driven by captured motion. In other words, the motion has been imposed to the virtual human. Then each pose from the motion is used by analysis functions. The Figure 2 shows the skeleton model (Top Right) in motion capture system and the virtual human (Top Left) in JACK<sup>®</sup> driven by the captured motion. For instance, for each pose, an OWPA (Ovaks Working Posture Analysis) analysis can be conducted (Bottom). The analysis system uses different colours to inform designers for example 'Yellow warning' and 'Red warning'.

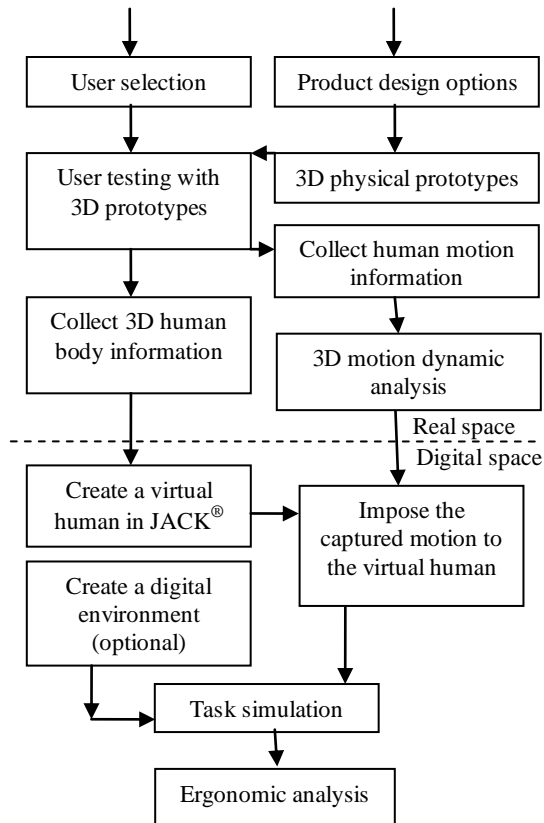


Figure 1. A method with virtual human models and imposed motions

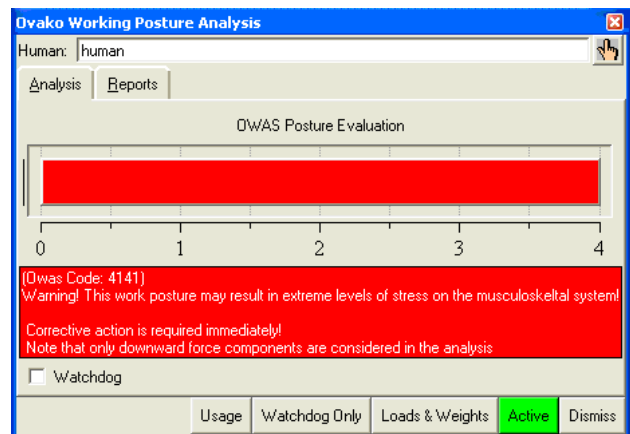
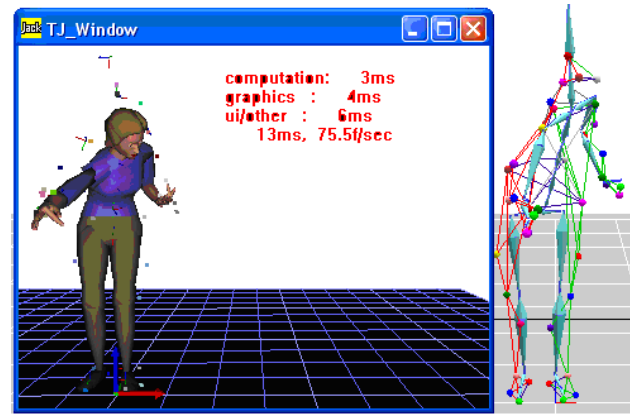


Figure 2. Virtual human with imposed motion for design studies

#### IV. CASE STUDY

The proposed method has been applied to a conceptual bed design for older people. The 3D body scanning technique has been used for collecting 3D human body information. We used a digital human with proposed method to adequately address bed design issues with ergonomic analysis. The bed design is targeted for older people to use at home for more independent living. The design issues include (1) how bed mattresses affect the user's comfort level and easiness of getting out of the bed, (2) whether an assistive device (haptic interface design) is necessary and what is its impacts on the user interaction, and (3) to what extent the older people are sensitive to these design options.

##### A. User selection

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

This study utilised three employees who were solicited on a voluntary basis. The participants in this study were selected on the basis of their age and body sizes (see Table 1). This selection criterion was chosen to ensure that each participant could represent a group of people. They were all in good physical condition. The performances of different aged participants can be used for comparison.

TABLE I. PARTICIPANTS

	Participant 1	Participant 2	Participant 3
height	178cm	166cm	175cm
Gender	male	female	female
age	21	38	61

### B. Motion capture design

The equipment for data collection included a motion capture system (Eagle Digital System, MotionAnalysis® Corp., Santa Rosa, CA), and a video camera to record the task process. Forty-one reflective markers were attached to the motion capture suite representing a typical marker set consistent with JACK® Motion Capture (Mocap) Toolkit (see Fig. 3). Seven infrared cameras were set up in our research laboratory and calibrated to obtain 3D coordinates of these markers at 60 frames per second. Cameras were clamped on the tripods above the test space (see Fig. 4).

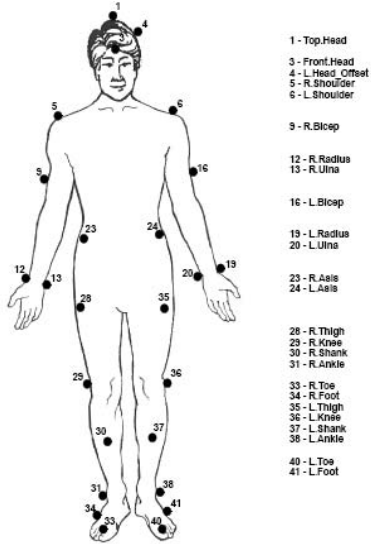


Figure 3. Forty-one marker template for JACK MoCap toolkit



Figure 4. Motion capture tests with different design options

### C. 3D bed design options and prototypes

The design options include the use of soft mattresses and hard mattress and the use of single and two handles.

A bed with a soft mattress is shown in Figure 4. A bed with a flat wood plate (similar to a hard mattress) is not shown. The soft bed was then tested with a single assistive handle on one side and two handles on both sides. These four design options are illustrated in Figure 5. Options A and B were aimed to test the impacts of stiffness and softness of the mattresses rising up from a bed in terms of easiness and comfort. Options C and D were designed to identify (1) whether a handle on a side is helpful for rising from a bed and makes rising easier and (2) whether two handles would be better than one.

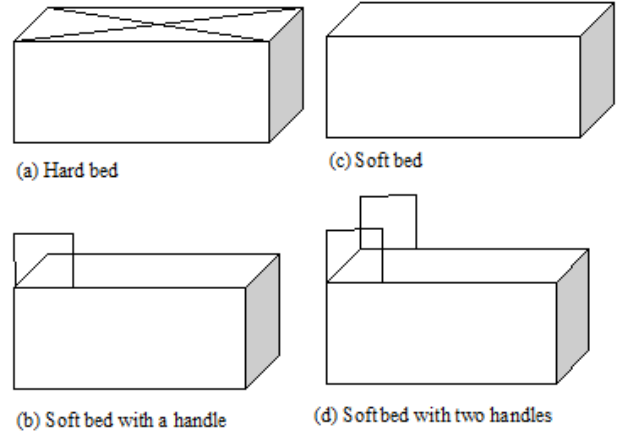


Figure 5. Design options and 3D prototypes

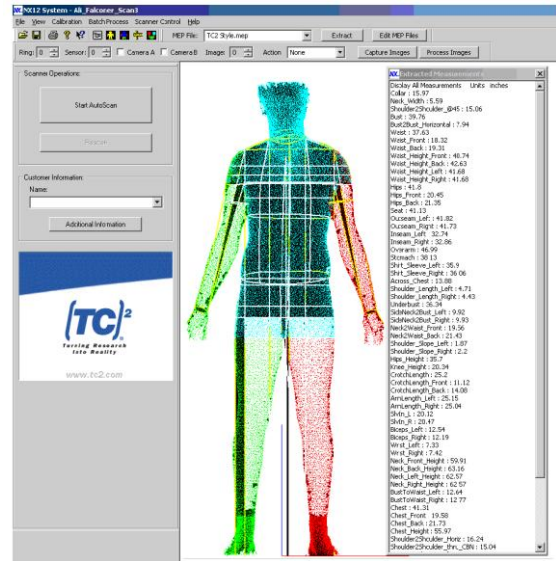


Figure 6. Collection of human body information with a 3D body scanner

### D. Collecting body information by body scanning

In order to best match a virtual human figure in JACK® with a real user, the real user's 3D body was first scanned by a 3D body scanner NX12 [TC]² and then relevant dimensions were extracted for use in JACK®. This body scanner can scan a real human in a standing position as shown in Figure 6 (left). The purpose of the body scanning technology is for textile and clothing design. The [TC]² 3D Body Scanner can scans the whole

body in less than 6 seconds and rapidly produces a true-to-scale 3D body model. The 3D body scanner features greater accuracy than manual measurements. After the scanning, 26 advanced scaling factors can be extracted and applied for updating the virtual human in JACK®. Although in this case study we used a 3D body scanning system, the proposed method does not exclude the use of alternative method to measure body sizes of a real user.

#### E. Experimental procedure

The experimental procedure was as follows. First, the participants wore the motion capture suit and then had their body scanned. After that, the participants would test the hard bed and soft bed, followed by the soft bed tests with either a single handle or two handles. The task included getting into the bed, facing upwards and rising from the bed. For each test, the participants would have two experimental trials. The motions during the test were then recorded. The way of getting into bed and rising from it is free. After all the motion capture tests were completed, we carried out post-processing tasks. Firstly, we extracted relevant body dimensions from scanned body data for the participants. Secondly, we post processed captured motion data by removing noises and interpolating new positions for non-visible markers temporarily blocked by body parts from cameras.

Finally, we used JACK® to ergonomics studies. In JACK®, firstly, a virtual human model with specified gender, height, weight and percentile parameters was created from a human library. Afterwards, it was modified by using body part scaling functions to make the virtual human match the real design participant and then refined based on the scanned body part data. Note that if the virtual human is not good enough to match the real human, her skeleton model with joint positions would be quite different, as a result, the motions applied to each joint at a time would produce very bad simulation effects. Secondly, captured and post-processed motions were played in the motion analysis software EVaRT and these motions were imposed on the virtual human in JACK® to conduct ergonomic analysis. In the EVaRT, a skeleton model was generated through creating a “SIMM Orthtrak Model” with a joint template and an initial T-pose motion. The peer-to-peer streaming communication was enabled through its NIC address (the machine’s IP address) on the EVaRT side. On the JACK® side, the ‘Motion Capture’ module with the selection of ‘Motion Analysis’ was set up and the communication with the EVaRT was enabled by specifying the IP addresses for JACK® and Motion Analysis applications respectively. Connection between the Motion Analysis subject and the virtual human in JACK® was then established. After setting up these connections, the captured motions were used to constrain the virtual human movements. When playing back a captured motion clip in EVaRT, the virtual human in JACK® will be imposed with the motions from EVaRT. Finally, task analysis modules were used to do task analysis.

#### F. Armrest test

We hypothesized that two armrests on both sides of a bed may perform better. In order to test our hypothesis, we studied Ovako Working Posture Analysis (OWPA) for a single armrest and two armrests. The initial test we carried out was to determine whether two armrests for the bed are better than one armrest.

For OWPA analysis, the JACK® software system can give different evaluations for various postures. The feedback from the system is graphically displayed in different colors to make it easier to understand and associated with text information as shown in Fig 7.

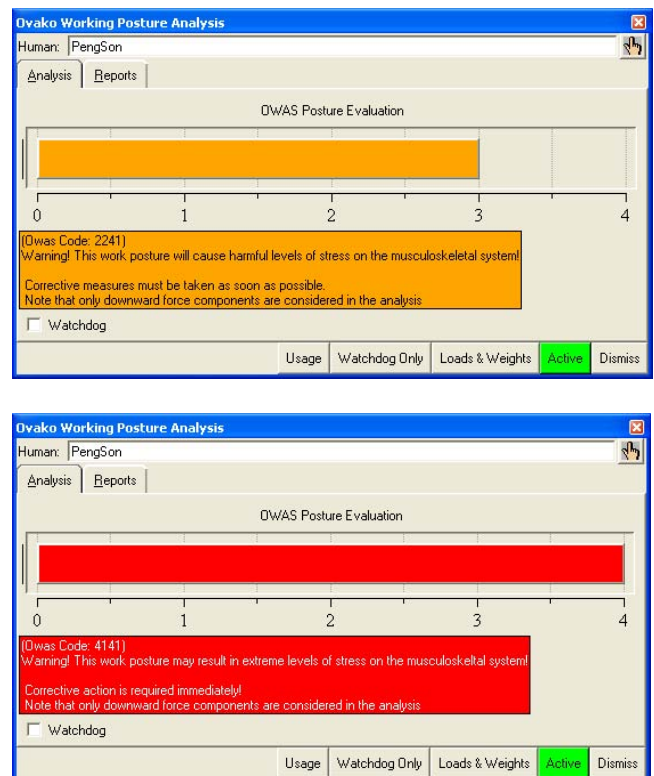


Figure 7. OWPA posture evaluation information. The former (top) warned a harmful posture (in yellow or orange colors) while the later(bottom) warned a posture in extreme levels in red color.

From the test results, it can be seen that a bed with one support is easier to use than with two supports. It is different from our initial assumption. It could be explained as the participants used one hand on one support meanwhile the other hand used the bed surface for assistance. This allows more natural postures in action.

#### V. CONCLUSIONS AND DISCUSSIONS

This paper presented a new method for using digital humans with imposed motions and motion dynamics to evaluate customised product design. We have demonstrated the use of a 3D body scanner for accurately measuring 3D body dimensions of real users and the use of a 3D motion capture system to obtain real motions (postures) of design interaction. Once the link between the JACK® ergonomic assessment tool and the motion capture

too is well established, the ergonomic analysis in JACK® can provide useful design evaluations graphically.

The proposed method can perform ergonomic assessments by a digital human with real task motion data which can reduce the deficiency of predicting task postures by engineer or ergonomists. The method combines the advantages of both real task motion and digital human simulation. With this method, the task motion is real instead of synthetic one. In ergonomic assessment software such as JACK®, there is no need for creating a CAD-based virtual product model to generate synthetic motions because a digital human with imposed real motion is enough to perform an ergonomic assessment. This method can support ergonomic evaluations of product design at the conceptual stage without its CAD models of the product.

#### REFERENCES

- [1] D. B. Chaffin, Digital human modelling for vehicle and workplace design, Warrendale, PA: Society of Automotive Engineers. 2001.
- [2] J. Z. Yang, J. H. Kim, K. Abdel-Malek, T. Merler, S. Beck, and G. R. Kopp, "A new digital human environment and assessment of vehicle interior design", *Computer-Aided Design*, Vol. 39, pp 548-558, 2007.
- [3] U. Jayaram, S. Jayaram, I. Shaikh, Y. J. Kim, and C. Palmer, "Introducing quantitative analysis methods into virtual environments for real-time and continuous ergonomic evaluations," *Computers In Industry*, Vol. 57, pp 283-296, 2006.
- [4] N. Badler, C.A. Erignac, and Y. Liu, "Humans for validating maintenance procedures". *Communications of the ACM*, Vol. 45, pp. 57- 63, 2002.
- [5] R. Feyen, Y. Liu, D. B. Chaffin, G. Jimmerson, B. Joseph, "Computer-aided ergonomics: a case study of incorporating ergonomics analyses into workplace design," *Applied Ergonomics* Vol. 31 (1), pp.291-300, 2000.
- [6] D. B. Chaffin, "Human Motion Simulation for Vehicle and Workplace Design," *Human Factors and Ergonomics in Manufacturing*, Vol. 17(5), pp.475-484, 2007.
- [7] C.F. Kuo, C.H. Chu, "An online ergonomic evaluator for 3D product design," *Computers in Industry*, Vol. 56(5), pp.479-492, 2005.
- [8] J.Y. Du, and V. G. Duffy, "A methodology for assessing industrial workstations using optical motion capture integrated with digital human models," *Occupational Ergonomics* Vol. 7, pp 11-25, 2007.
- [9] K. Li, V. G. Duffy, and L. Zheng, "Universal accessibility assessments through virtual interactive design," *Human Factors Modelling and Simulation*, Vol. 1(1), pp. 52-68, 2006.
- [10] S. W. Chang, and M. J. J. Wang, "Digital human modelling and workplace evaluation: using an automobile assembly task as an example," *Human Factors and Ergonomics in Manufacturing* Vol. 17(5), pp 445-455, 2007.
- [11] Michael Gleicher, *Retargetting motion to new characters*, Proceedings of the 25th annual conference on Computer graphics and interactive techniques (1998), pp.33-42.
- [12] S.F. Qin, D.K. Wright, J.S. Kang, and P.A. Prieto, "Use of 3d body motion to freeform surface design," *Proc. of IMechE, Part B, Journal of Engineering Manufacture*, Vol. 220(2), pp. 335-339, 2006.
- [13] S.F. Qin and D.K. Wright, "Progressive Surface Modelling Scheme from Unorganised Curves," *Computer-Aided Design*, Vol. 38(10), pp. 1113-1122, 2006.
- [14] X. Yi, S.F. Qin, and J.S. Kang, "Generating 3D architectural models based on hand motion and gesture," *Computers in Industry*, Vol. 60, pp. 677-685, 2009.
- [15] S. Paquette, "3D Scanning in Apparel Design and Human Engineering," *IEEE Computer Graphics and Applications*, Vol. 16(5), pp. 11-15, 1996.
- [16] C. L. Istook, S.J. Hwang, "3D body scanning systems with application to the apparel industry," *Journal of Fashion Marketing and Management*, Vol. 5(2), pp.120-132, 2001.



# *Volume Deformation Based on Model-Fitting Surface Extraction*

Qian Xu, Duke Gledhill, Zhijie Xu

CGIV Research Group, University of Huddersfield  
Huddersfield, West Yorkshire, United Kingdom

[q.xu@hud.ac.uk](mailto:q.xu@hud.ac.uk), [d.gledhill@hud.ac.uk](mailto:d.gledhill@hud.ac.uk), [z.xu@hud.ac.uk](mailto:z.xu@hud.ac.uk)

**Abstract** - Over the last decade, visualization techniques for 3-dimensional volumetric models, especially those that can be performed on PC hardware platforms, have attracted intensive attention in the research communities. The rapid evolution on PC computers, specialist hardware, and even gaming consoles have accelerated this trend and seen the volume model-based applications being greatly extended from industrial design, medical simulation, to entertainment usage and beyond. As part of the effort, the interactive manipulation of the appearance of volume models, often referred as volume deformation, has become a research hot-spot due to its potentials in revealing the models' internal structures and material characteristics. This paper reports an innovative volume deformation method based on a self-extracting mechanism for the so-called "control lattice" from the "surface" of a volume model, which can then be applied on the entire model or a specifically segmented part of it based on user requirements. The detail level of the extracted control lattice can be customized based on Active Surface algorithms for ensuring the interactive rate and the final resolution for a particular application.

**Keywords**- *Volumetric Model; Volume Deformation; Control Lattice; Adaptive Segmentation; Active Surface*

## I. INTRODUCTION

For showing the internal and implicit information of a 3-dimensional (3D) volumetric model, various PC-based visualization techniques have been developed in the past decade, such as Ray Casting and Splatting. However, for applications such as medical operation planning and design function simulations, volume models need to be "operable" in a rigid manner like splitting or slicing, as well as in a non-rigid form such as elastic deformation.

Volume data sets are often obtained from special industrial cameras and medical scanners, or being generated by mathematical models and algorithms. These models do not come with a mesh/polygonal representation of the compositing voxels, a synonym for volumetric-pixels, nor containing any intrinsic topological information of the models' internal structures and "material" properties - if applicable.

Most of the popular volume visualization techniques performed on PC-grade hardware assemble 3D models through accumulating the voxel contributions to the final "pixel" colours on the 2D virtual image plane. Although first proposed in the 1980s, volume visualization and its applications had really started gathering momentum since the late 1990s attributing to the ever more "powerful" PC capabilities. Volume deformation, as a branch of the trend,

is stemmed from the visualization progression and had been focusing on the manipulation of volume models in a pre-defined or free-form manner and their corresponding control mechanisms.

Generally speaking, volume deformation approaches can be classified into two groups: rigid and non-rigid; while the prior can be considered as an extension of the Computational Solid Geometry (CSG) model with its mathematical representations defined by works such as Computational Volume Geometry (CVG) [1]. The latter approach preserves the underlying volumetric assemblies while enabling the description of the physical constraints and relationships among those elements. In the actual volume deformation applications, this approach can really meet the requirements of FFD (Free Form Deformation) which rigid ones cannot easily achieve [2, 3].

Unlike rigid deformation, the non-rigid approaches mostly require a control lattice (analogy to a web of "control" vertices) to be created for each operation before any displacement computations can start. For example, DOGME (Deformation of Geometric Models Editor) method in physically-based simulations, which is a constrained-based method, generally assumes a lattice which is the finest choice and passes the computed displacements to the underlying model or elements in the form of polygonal topologies [4]. For achieving the special visualization objectives, few non-rigid approaches totally rely on the dedicated rendering methods with corresponding calculation and interpolation works. For example, ray-deflectors-based methods, which do not move any voxels, manage to exhibit the deformed results by changing the visualization properties [5]. Based on the various requirements of the precision and usage in actual applications, the non-rigid approaches can be further classified into physically-based and empirical methods [6].

In the applications of physically-based volume deformation, a control lattice can be of a 2D planner form, or a 3D cubic or cylindrical form, or even being parametrically defined, which can represent material properties of the sampled voxels [7, 8]. However, due to the complex process often involved in the construction of a control lattice, and the shape and tessellation of it, the overall quality of the deformation varies.

In the surveyed volume deformation pilots, a small difference in the lattice definition strategy can lead to substantially varied voxel displacement results [9, 10]. Generally, a volume model is defined without any intermediate boundary representation. Westerman and Ertl



have developed a method for texture-based rendering of volume data based on a uniform regular grid as well as over a tetrahedral grid, which partially resolved the problem [11]. However, these techniques are focused on rendering aspects rather than interactive manipulations and precise simulations.

An alternative approach is to extract an intermediate iso-surface (polygonal mesh) from the volume data set that can be used for further manipulation and processing, such as collision detection, shadow casting, and animation. The most commonly used algorithms for iso-surface extraction are mainly derivatives of the Marching Cubes (MC) algorithm developed by Lorensen and Cline [12], which construct the iso-surface generation mechanism following the renowned “15 unique cube configurations” [13].

In this project, an innovative volume deformation approach is proposed based on an embedded MC process for constructing the so-called “model-fitting” control lattices. This paper will cover the Snake-based volume data orientation and the assistant design for MC-based lattice extraction in Section III and IV. The design and implementation of GPU-based Octree data structure are introduced in Section V with the deformed results.

## II. SYSTEM DESIGN

In this research, a volume deformation pipeline based on the latest programmable GPUs has been developed as shown in Fig. 1. This paper mainly covers three sections: Data Segmentation, Mesh Modification and Spatial Relationships Determination.

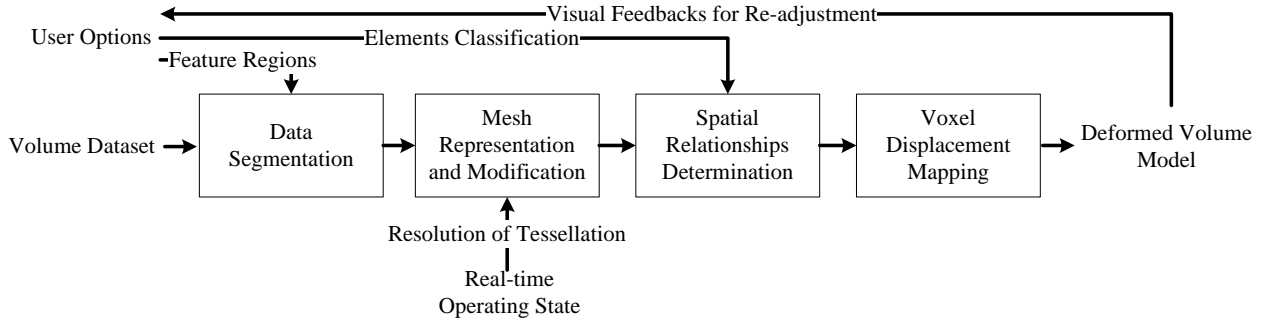


Figure 1. The pipeline of improved construction of lattice for volume deformation.

### A. Data Segmentation Process

For reducing the workload of processing the entire data set in each processing cycle, a segmentation process has been designed to detect and track the “interested” objects which are mixed with other useless parts, for example, an organ in a MRI-scanned human body, or a section of the generated “point clouds”. This step will reduce the overall memory “footprint” for storing and accessing the volume data set at runtime, and the complexity of the iso-surface extracted for forming the control lattice. Besides, implementing deformations on respective classified data sets can really support a shape-changed volume model containing different grades of

transformation in the physically-based deformation applications.

### B. Mesh Representation and Modification Process

However, the classified data set still leads to a time-consuming iso-surface extraction on CPUs for MC mechanism. The system developed in this project applied a GPU-accelerated (Graphics Process Unit) iso-surface extraction method for alleviating this problem. Both CPU-based and GPU-based MC processes often make extracted surfaces consist of far too many vertices to be used as a control lattice, which is literally a large spring network. The prototype system contained an adaptive tessellation engine for adjusting the resolution of the lattice via changing the number of vertices on the iso-surface.

### C. Spatial Relationships Determination Process

Similar to the DOGME method in traditional vertex-based applications, this project is delivering the displacement on a control point to underlying elements which are “preformed” by many masses in a designed mass-spring model. A dedicated Octree data structure was constructed to manage the mass partition stage and determine the internal relationships between voxels. By implementing this data structure on GPU, the mass accessing and the internal spring-like relationships can be more efficiently implemented.

## III. VOLUME MODEL SEGMENTATION

Separating objects from backgrounds or from each other through segmentation, which is an important process in computer vision and image processing, can reduce the complexity of the subject studied with its applications often found in measuring object size, tracking vehicle movement and scientific visualization. The volume data segmentation operation devised in this project is mainly based on the Active Surface that is a 3D extension of Active Contour Theory.

Active Contour (or Snake) is a method that enables delineating 2D outlines from an image. This technique contains a “spline” which follows the energy minimization rule and can be deformed by “forces”. The forces are determined by an assembly of intrinsic and extrinsic constraints. The intrinsic constraints are

determined by the material properties of the spline, such as mass distribution parameter and viscosity of neighbouring medium. The extrinsic constraints represent the external forces which links splines (lines or surfaces) to underlying elements (pixels or vertices). The Energy  $E(V)$  can be calculated according to “forces”. Kass et al. proposed the equation of Energy Minimization in 1988 [14]:

$$E(V) = \int_0^1 (E_{\text{internal}}(V) + E_{\text{external}}(V)) dV \quad (1)$$

where  $E_{\text{internal}}$  is the internal energy of the bended spline and  $E_{\text{external}}$  serves as external energy acting on the spline.  $E_{\text{external}}$  contains  $E_{\text{image}}$  represents the image force acting on the spline and  $E_{\text{con}}$  denotes external constraint forces defined by the user.

Active Surface is a 3D variation of the Active Contour technique. In this project, region-based Active Surface algorithm is used to analyze 3D data sets. Its mathematical model can be represented by a parameterized surface  $\delta$  [15]:

$$\delta : \Omega^2 \rightarrow \mathbb{R}^3 \quad S(U, V) = [x(U, V), y(U, V), z(U, V)]^\delta \quad (2)$$

Where  $(U, V)$  determines the coming changes of surface  $\delta$ . And the changed surface  $S(U, V)$  can be represented via an assembly of moved vertices. The calculation of minimal energy in the sampling region  $\delta$  is represented via equation:

$$E[\delta] = \phi E_{\text{smooth}} + (1 - \phi) E_{\text{region}} \quad (3)$$

where  $\phi$  is a pre-input parameter that weights the significance of smoothness term. For processing 3D data, Mille proposed the smoothness energy for calculating 3D data sets:

$$E_{\text{smooth}}[\delta] = \iint_{\Omega^2} \left\| \frac{\partial S}{\partial U} \right\|^2 + \left\| \frac{\partial S}{\partial V} \right\|^2 dU dV \quad (4)$$

Surface  $\delta$  separates the object into an internal domain  $\gamma_{\text{in}}$  and an external domain  $\gamma_{\text{out}}$ . Based on the Chan-Vese model [16], the energy of 3D domain can be calculated in

$$E_{\text{region}}[\delta] = \iint_{\gamma_{\text{in}}} \mu_{\text{in}} dU dV - \iint_{\gamma_{\text{out}}} \mu_{\text{out}} dU dV \quad (5)$$

where  $\mu_{\text{in}}$  and  $\mu_{\text{out}}$  are intensity descriptors inside and outside the surface respectively.

At the beginning of the Active Surface segmentation process, there is a step for defining the “3D splines”. Fig. 2  $A_0$  and  $B_0$  illustrate the definition of rectangle-based splines in any two cross sections in the volume model. Based on the collection of 2D splines (Fig. 2  $A_1$ - $A_5$  and  $B_1$ - $B_5$ ), the region-based Active Surface algorithm can be implemented for direct and continuous analysis of 3D data sets. Its results are point-based and processed in the following MC stage.

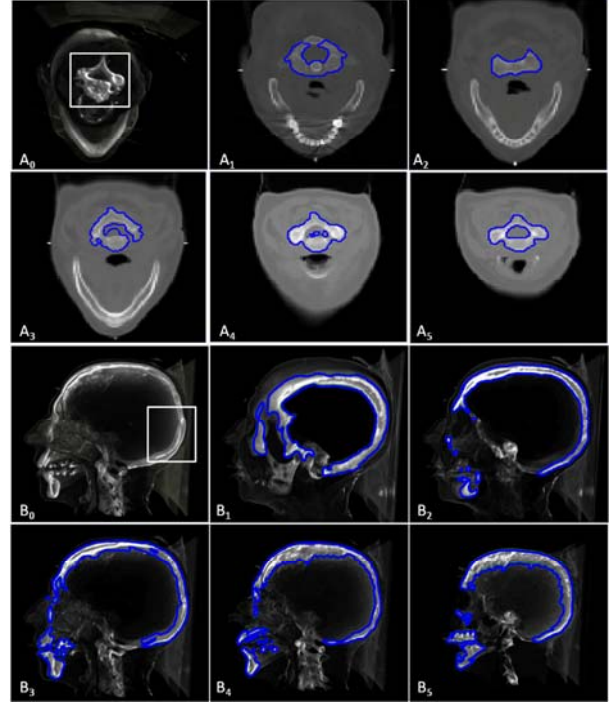


Figure 2. Image ( $A_0$ ) and ( $B_0$ ) respectively represent the ways of determining interesting-domains in axial and sagittal cross sections. Image ( $A_1$ - $A_5$ ,  $B_1$ - $B_5$ ) are snapshots of the detected regions.

#### IV. CONTROL LATTICE GENERATION

##### A. Marching Cubes and Iso-surface Extraction

In this project, MC is used to create the so-called “model-fitting” control lattice based on the iso-surface extracted from the volume data set. It travels through the volume model contained in an imaginary cube consisted of numerous “cells” as shown in Fig. 3 (a) and (b). The MC algorithm tests each voxel and produces vertices within the corresponding cells. The density values at the 8 vertices of a cell are evaluated based on the “signs” determined by their relative positions to the iso-surface. The voxels that lie on the boundary between the model and the “empty space” (voxels with null values) are then processed for generating polygons according to the standard “15 unique cube configurations” as shown in Fig. 3 (c).

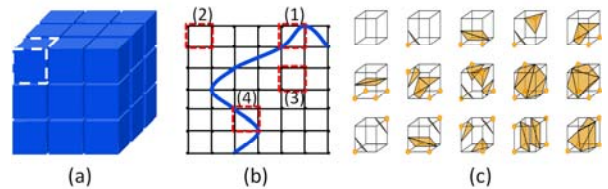


Figure 3. (a) represents the “imaginary cube” and (b) shows different conditions that happened in the sampling progress. (c) illustrates the standard “15 unique cube configurations.”

Based on these fifteen basic configurations, there will be a total of 241 derived forms. It is difficult to make a definite description of these 256 cases without labelling the eight vertices. As a result, the tri-linear interpolation (checking the sign for inside of the cube) is used in this

project to avoid ambiguity. As stated by Engel [Engel et al, 2004], one of the advantages of the Marching Cubes algorithm is its locality - the voxels are processed one-by-one based on local information only. The computational processes for MC can be readily parallelized and “mapped” on GPUs for harnessing its data parallelism. Fig. 4 shows the output iso-surfaces from the segmented data sets in the devised program in the form of a closed polygonal mesh.

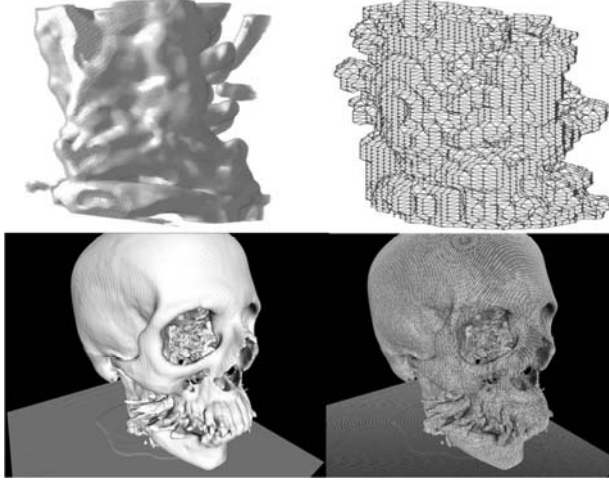


Figure 4. These isosurfaces are respectively extracted from the segmented throat data and skull data in a MRI-scanned human head.

### B. GPU-based Adaptive Control Lattice

The surface extracted from the MC process contains a huge number of polygons, e.g. there are 2 million vertices in the iso-surface-based skull model in Fig. 4. This complicated mesh can be slow to interact with when used for calculating deformation parameters such as vertex and voxel displacement offsets and rotational angles. Through controlling the tessellation of the extracted iso-surface, an adaptive control lattice can be formed.

There are existing GPU models supporting hardware-driven tessellation operations that can be adopted for this purpose, e.g. GPU-based Catmull-Clark subdivision is available in mesh-based deformation for adapting the density of vertices in the deforming regions [17, 18]. It is used to implement the real-time distribution of vertices during the deformation, i.e. carry out LOD-based control for managing the number of vertices. However, in this research, the choice of data structure, derived operations and the consideration of its’ familiar artefacts called T-junction determine that Catmull-Clark subdivision needs to be replaced by a triangle-based subdivision scheme – Loop Subdivision [19]. Its reverse scheme is chosen to achieve the More-to-Less operation for reducing the vertex count and implementing vertex management in an “on-the-fly” style.

The (reverse-) Catmull-Clark subdivision and (reverse-) Loop subdivision schemes are both based on the management of object vertices and their neighbouring points. Accordingly, implementing reverse Loop subdivision on GPU can be technically divided into three main steps. Firstly, the object mesh is separated into a

series of sub-meshes according to the results of triangular partition (represented by the blue triangles in Fig. 5 Step 1). This process does supply outstanding vertices for the second step to find their surrounding points which share the same lines with the objective vertices. After this finding process, the second step carries on labelling all participant vertices for GPU-based vertex storage in the third step. For all vertices stored in texture memory in the form of 2D arrays with the default index numbers ( $\lambda$  in Fig. 5 Step3), the single-threading-management of vertices in the CPU-based Loop subdivision can be implemented more efficiently via the GPU-based multithreading-operations that support synchronized vertex-release processes and the real-time combination of processed arrays. For conveniently understanding the implementation of GPU-based Loop subdivision in this project, the schematic meshes are all ideal and the valence of each vertex is 6. The sketch maps of each reverse subdivision process and simplified results are all respectively illustrated in Fig. 6 and 7. The polygon account is hugely reduced from millions to thousands, e.g. from 2.4 million vertices to 12 thousands in the skull model.

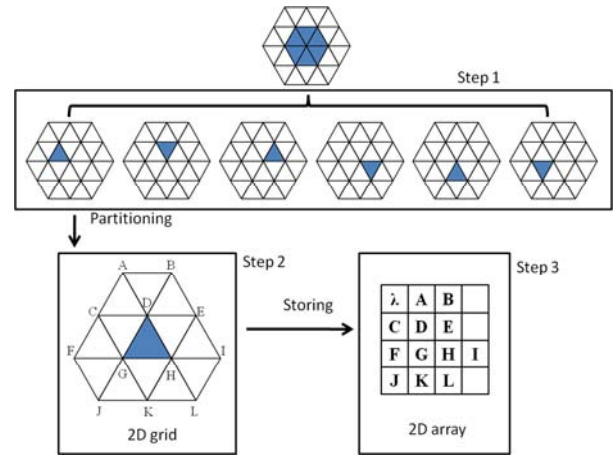


Figure 5. Illustrations for GPU-based reverse Loop subdivision.

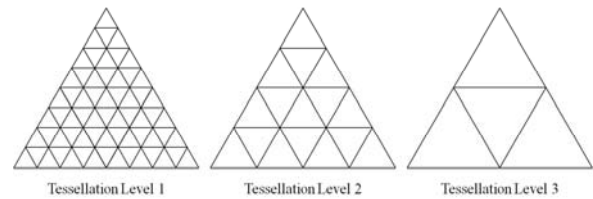
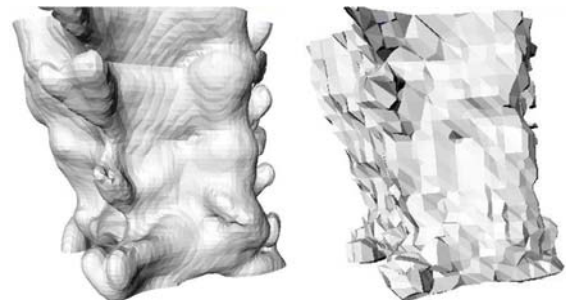


Figure 6. Continuous construction of a triangle primitive with decreasing tessellation levels.





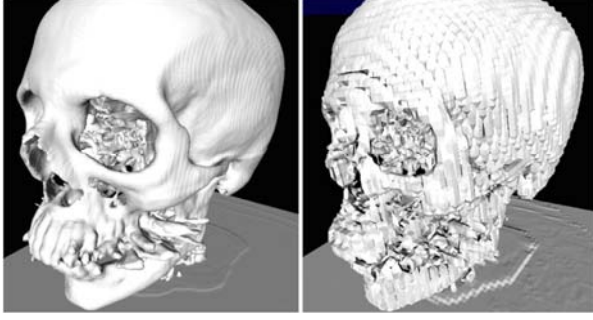


Figure 7. The illustrations of adaptive MC process in different tessellation levels.

## V. VOLUME DEFORMATION

After finishing the model-fitting lattice construction, this research uses octree data structure to separate the volume data set for implementing voxel-based masses and parameterized elastic characteristics in the designed mass-spring model. The implementation work mainly consists of three processes. First one is “encoding” the volume data set for GPU-based storing process. Similar to other applications of GPU-based 3D octree data structure, the whole volume model is equally divided into an assembly of elements which generally contain  $2^N \times 2^N \times 2^N$  voxel(s) ( $N$  is the depth of octree), and the partitioned results are stored in an 8-bit RGBA 3D texture in the form of  $N^3$ -tree. The coordinates of each subset or element are stored in RGB channels and the Alpha channels record the results of identifying a pointer to the content of a leaf or a child node. 0, 1 and 0.5 respectively indicate pointing to an end of the current branch (voxels with null values), an available child node and an indexable content.

After storing process, the second process is tracking the object node's related leaf contents or nodes on different levels (3D representations are shown in Fig. 8). In order to accomplish this goal, Tree-Lookup function, which is a common module in various applications of GPU-based Octree data structure, determines the resulting contents or nodes based on the choice of Alpha value of each element on the current level. The results are a unit cube or a cubic region with mathematical expression  $[O_0, O_1]^3$ .

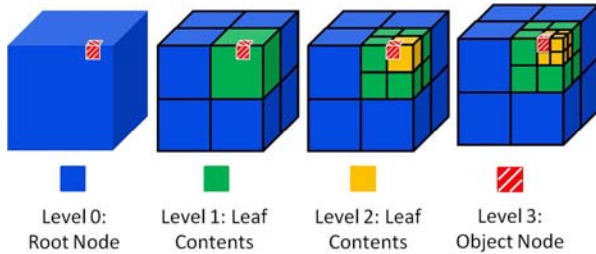


Figure 8. Octree-based method for tracking the objective area or element (represented by the yellow or red square).

The 3<sup>rd</sup> process uses these detected contents or nodes to represent different grades of transformation in the physically-based deformation of segmented parts. In this project, the mechanism is proportionally assigned different scales of original displacement (contains distance

and direction) to corresponding contents or nodes. The related outcomes are an assembly of spatial coordinates. In Fig. 9, the different grades of deformation are carried out via decreasing the displacement along the track (dashed) of the moved control point (solid point). Meanwhile, a simple indicator is added to survey the points value to testify that the fixed number of points ensure no manmade or interpolated artefacts.

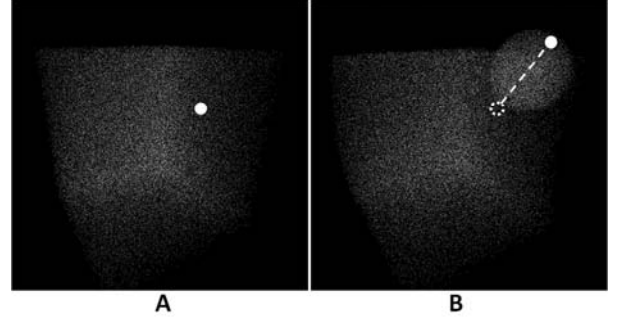


Figure 9. A point-based representation of a volume model (A); the control point (solid point in (A)) obtains a displacement and the deformed result is shown in (B).

The current system is designed and constructed using MATLAB, OpenGL, and CUDA APIs under the Visual Studio/VC++ programming environment. A desktop PC with an Intel Core2 2.40GHz processor and an NVIDIA GeForce GTX260 GPU with 2GB memory have been used for experiments. Because no new voxels are created, the following work is directly following the initial spatial arrangement rule to “filling” original voxels to the new assembly of coordinates extracted from the Octree data structure. Fig. 10 has shown the corresponding result of Fig. 9 in (a) and other derived deformation outputs (reversed and multiple) are shown in (b) and (c).

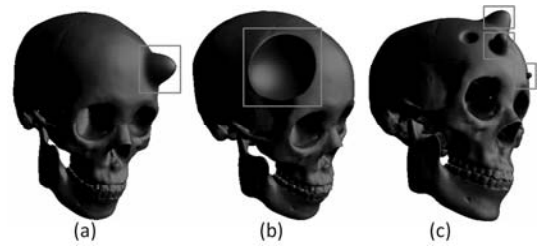


Figure 10. The appearance of various skull deformations.

In Fig. 11, these vertical-sections illustrate the deformed internal/external regions in a human skull model. The effects of physically-based deformation in my project, which need to represent the details of volume deformations with different distributions of displacement in different materials, had been highlighted by the white rectangles.

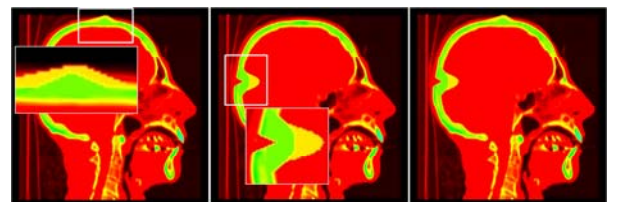


Figure 11. Lower line illustrates the corresponding vertical-sections.

And the frame-rates are totally distributed between 49 and 67 (higher than 30 fps required in actual applications). Therefore, these records can demonstrate the feasibilities of interactive manipulations and multiple deformations in my physically-based volume deformation.

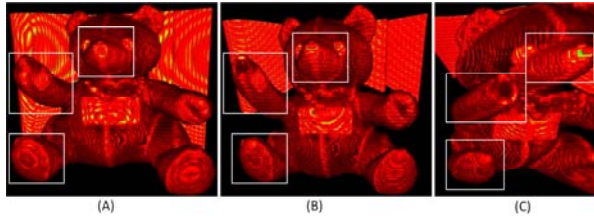


Figure 12. Texture-operation-based volume deformation.

## VI. CONCLUSIONS

In Fig. 12, B and C show two deformed teddy-bears which are generated via stretching the initial model (A) in two directions. This deformation method belongs to the texture operation-based volume deformation which is one branch of the non-physically-simulated applications. Its mechanism is moving the carriers to influence the corresponding elements, i.e. modifying textures to transfer voxels. In current volume deformation applications, non-physically-based methods are barely used for simple displaying, colourful rendering or conceptual expiations. In order to achieve imprecise simulations, they must require helps from the complicated segmentation algorithms to carry out certain texture pre-processing for fixing its lack of element-level-operation. Consequently, these designs always suffer from the time-consuming processes, resulting artefacts in texture pre-processing and low interactive rates.

The novel physically-based volume deformation pipeline developed in my project can manage the basic elements (voxels) for achieving complicated simulations and related requirements. It has replaced the conventional methods for constructing control lattices using an improved model-fitting MC-extracted surface mesh. The latest GPU tessellation capacity adopted in this research has further enabled hardware-driven complexity control that optimized the frame of the extracted mesh. Experiments carried out in the research have shown improvements on the accuracy and flexibility of deformation types that can be performed using this design. It is anticipated that applications such as medical operations, simulations, industrial designs, and even computer games can benefit from this innovative volume deformation approach. Future work will focus on further improvements for system integration and real-time performance.

## ACKNOWLEDGMENT

I would firstly like to thank my first supervisor Dr Zhijie Xu for his great supervision and guidance during this research. I am deeply impressed by his erudition, and knowledge and attitude to science which inspired me to

keep on going. I would also like to show my grateful appreciation to Dr Duke Gledhill for his great help to me.

## REFERENCES

- [1] M. Chen and J. V. Tucker, "Constructive Volume Geometry," In *Computer Graphics Forum*. Vol. 19, No. 4, pp. 281-293, 2000.
- [2] N. Chen, R. Alterovitz, D. Ritchie, L. Cho, K. K. Hauser, K. Goldberg, J. R. Shewchuk and J. F. O'Brien, "Interactive Simulation of Surgical Needle Insertion and Steering," In *ACM Transactions on Graphics*. Vol. 28, No. 3, pp. 1-10, 2009.
- [3] C. Corea, D. Silver and M. Chen, "Feature Aligned Volume Manipulation for Illustration and Visualization," In *IEEE Transactions on Visualization and Computer Graphics*. Vol. 12, No. 5, pp. 1069-1076, 2006.
- [4] B. Sarvage, S. Hahmann, G. P. Bonneau and G. Elber, "Detail Preserving Deformation of B-spline Surface with Volume Constraint," In *Computer Aided Geometric Design*. Vol. 25, No. 8, pp. 678-696, 2008.
- [5] Y. Kurzion and R. Yagel, "Interactive Space Deformation with Hardware-assisted Rendering," *IEEE Computer Graphics and Applications*. Vol. 14, No. 5, pp. 66-77, 1997.
- [6] C. D. Corea, D. Silver and M. Chen, "Constrained Illustrative Volume Deformation," In *Computers & Graphics*. Vol. 34, No. 4, pp. 370-377, 2010.
- [7] M. Hong, S. Jung, M. Choi and S. W. J. Welch, "Fast Volume Preservation for a Mass-Spring System," In *IEEE Computer Graphics and applications*. Vol. 26, No. 5, pp. 83-91, 2006.
- [8] G. Yin, Y. Li, J. Zhang and J. Ni, "Soft Tissue Modeling Using Tetrahedron Finite Element Method in Surgery Simulation," *Proceedings of the 2009 First IEEE International Conference on Information Science and Engineering*. Nanjing, China, pp. 3705-3708, 2009.
- [9] R. Westermann and C. Rezk-Salama, "Real-time Volume Deformations," In *Computer Graphics Forum*. Vol. 20, No. 3, pp. 443-451, 2001.
- [10] S. Walton and M. Jones, "Volume Wires: A framework for Empirical Non-linear deformation of Volumetric Data Sets". In *Journal of WSCG*. Vol. 14, No. 3, pp. 81-88, 2006.
- [11] R. Westermann and T. Ertl, "Efficiently Using Graphics Hardware in Volume Rendering Applications," *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. Orlando, USA, pp. 169-177, 1998.
- [12] W. Lorensen and H. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," In *ACM SIGGRAPH Computer Graphics*. Vol. 21, No. 4, pp. 163-169, 1987.
- [13] K. Engel, M. Hadwiger, J. M. Kniss, A. E. Lefohn, C. R. Salama and D. Weiskopf, "Course Notes 28 II: Real-Time Volume Graphics," *ACM SIGGRAPH 2004 Course Notes*, pp. 1-282, 2004.
- [14] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models," In *International Journal of Computer Vision*. Vol. 1, No. 4, pp. 321-331, 1988.
- [15] J. Mille, "Narrow Band Region-based Active Contours and Surfaces for 2D and 3D Segmentation," In *Computer Vision and Image Understanding*. Vol. 113, No. 9, pp. 946-965, 2009.
- [16] T. Chan and L. Vese, "Active Contours without Edges," In *IEEE Transactions on Image Processing*. Vol. 10, No. 2, pp. 266-277, 2001.
- [17] E. Catmull and J. Clark, "Recursively Generated B-Spline Surfaces on Arbitrary Topology Meshes," In *Computer Aided Design*. Vol. 10, No. 6, pp. 350-355, 1978.
- [18] M. Pharr and R. Fernando, "GPU Gems 2: Chapter 7. Adaptive Tessellation of Subdivision Surfaces with Displacement Mapping," In *GPU Gems2: Programming Techniques for High-Performance Graphics and General-Purpose Computation*. Printed in USA, 2004.
- [19] C. Loop, "Smooth Subdivision Surfaces Based on Triangles," Thesis for the degree of Master of Science, 1987.

# Numerical Simulation of Natural Frequencies in the Design of Micro Air Vehicle Structures

Yanan Yu (\*), Carlo Ferri (\*\*), Qingping Yang (\*\*), and Xiangjun Wang (\*)

(\*)Tianjin University, State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin, 300072, China

(\*\*)Brunel University, School of Engineering and Design, AMEE, Uxbridge UB8 3PH, UK

yanan.yu@brunel.ac.uk, carlo.ferri@brunel.ac.uk

**Abstract**—The mechanical resonance of structures is often blamed for their collapse. In the attempt to prevent such usually catastrophic events, the identification of the natural frequencies has usually become a constituent part of the design activity of mechanical structures. Micro Air Vehicles (MAVs) are small size unmanned aircraft. This study aims to give a place to the numerical finite element method (FEM) simulations of the natural frequency during the design stage of the MAVs. In particular, the effect of selected materials and geometrical parameters on the simulated first natural frequency has been analyzed in this study. Carbon fibre reinforced polymer (CFRP) appeared to confer the MAV a significant higher first natural frequency than the aluminum alloy investigated (AA2024). Also, given any of the material investigated (CFRP, aramidic fibre reinforced polymer, AA2024), it was found that MAVs simultaneously having a large radius and thickness of the outer shell and a small height have first natural frequency far larger than in any other geometrical configuration.

**Keywords**- numerical simulation; finite element method (FEM); natural frequencies of structures; micro air vehicle (MAV); exploratory data analysis (EDA)

## I. INTRODUCTION

Small-scale, automatic, unmanned air vehicles are often referred to as 'Micro Air Vehicles' (MAVs) [1] [2]. MAVs play a significant role in many fields of modern world like for example military surveillance, environment monitoring and scientific mapping [3]. Due to the small scale and the light weight MAVs are more sensitive to atmospheric disturbances (e.g. gusts) than other categories of aircraft (e.g. airliners). Assuring flight stability is therefore a prime concern in the design of MAVs [4]. For example, in [5] the stability characteristics and the control properties of hovering MAVs in wind gusts are described.

Beside flight stability, the identification of the mechanical resonant frequencies should also have a part in the process of designing a MAV. The natural frequencies and mode shapes of the dragonfly wing model were investigated by finite element method [6]. The declared intent of such a dragonfly study was to provide some help to the designers of aircraft of similar size. Like for any mechanical or civil structure (shafts, bridges, and buildings), the likelihood that a mechanical natural frequency of the structure becomes equal to that of the applied oscillating loads should be limited. Usually the resonance frequency of an aircraft depends on the scale of the object, its mass, aerodynamic shape, mechanical

structure and material [6]. The determination of the natural frequencies of circular cylindrical shells can play a part in the identification of natural frequencies of hovering MAVs. References [7] investigated a shallow spherical shell with large amplitude free vibrations. Reference [8] introduced a novel wave approach to study the vibration characteristics and predict the natural frequencies of circular cylindrical shells.

The prime aim of this study is to estimate the effects of selected design parameters (geometry and material) on the first natural frequency of an aero-elastic ducted-fan hovering MAV. Vertical take-off and landing and hover capabilities are among the main flight characteristics of hovering MAVs [9]. In Fig. 1, an example of a hovering MAV reported in [10] is displayed. The design parameters that significantly affect the natural frequencies of an aircraft structure should be identified during the design process. Knowing which design parameter has an effect on the natural frequencies would in fact enable the designer to make informed decisions to meet the natural frequencies functional requirements requested to the structure. A full factorial design of the numerical experiment is conducted in this study to estimate the effects of selected design parameters on the first natural frequency above mentioned. In Section II, the geometrical proprieties of the hovering MAV design considered in this study are presented. In Section III, the selected materials for the hovering MAV are introduced. In Section IV, the numerical results of the full factorial design for the numerical experiments are presented and discussed. The presentation and discussion method hinge on graphical



Figure 1. Example of a hovering MAV (from [10]).



representations typically used in exploratory data analysis (EDA). The conclusions are then drawn in Section V.

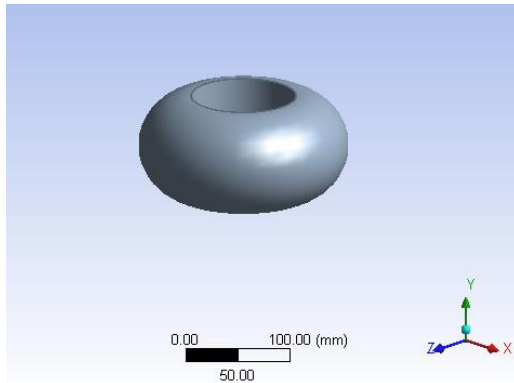
## II. GEOMETRY OF THE CONSIDERED HOVERING MAV

The aerodynamic shape of the aircraft contributes to determine the aircraft speed, the flight range and the flight stability. Hovering MAVs are typically made up of a shallow spherical shell and a vertical cylindrical duct. The geometry of the hovering MAV design examined in this study is shown in Fig. 2. Four dimensional parameters have been selected to investigate the effect that their values may have on the first natural frequency of the whole structure. For each of the four parameters two values were considered that span the range of values attributable to them without drastically change the overall form of the aircraft. Parameters and values are displayed in Table I.

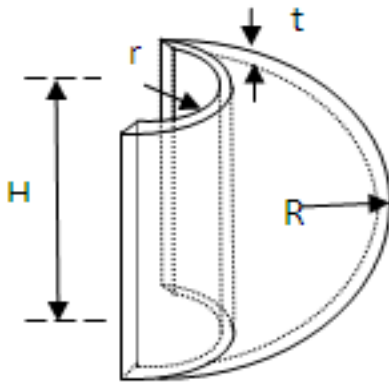
## III. SELECTION OF THE MATERIALS

The choice of an aircraft material significantly affects its flight performance. Usually, material and geometry are jointly selected to meet the aerodynamic requirements of the aircraft. In this study, an aluminum alloy and two different composite materials are considered for the hovering MAV design.

Aluminum alloys are typically used in aircraft



(2a) Hovering MAV geometry.



(2b) Geometrical parameters of interest.

Figure 2. The geometry of the hovering MAV. In (2a) the geometry of the whole aircraft is displayed, whereas in (2b) the investigated geometrical parameters are visible.

TABLE I. GEOMETRICAL PARAMETERS

Symbol	Description	Low / mm	High / mm
H	Height of the MAV	95	120
R	External radius of the shell	60	100
r	Internal radius of the cylindrical vertical duct	30	60
t	Thickness of the shell	5	10

structures and aerospace applications. Among these, aluminum alloy 2024 is one of the most used materials in aircraft structural components owing to its high strength-to-weight ratio. In particular, such an alloy is traditionally used for wings and fuselages [11].

Composite materials have been given large attention in the latest decades within the aerospace industry because of their high strength and their light weight. From a practical point of view, a composite material can be thought of as a multiphase mixture of at least two solids that are commonly referred to as the matrix and the reinforcement of the composite. In this study, two different composites are considered: an epoxy matrix reinforced with carbon fibre (i.e. carbon-fibre-reinforced polymer, alias CFRP) and an epoxy matrix reinforced with aramidic fibre (i.e. Aramidic-fibre reinforced polymer, alias AFRP). Aramidic fibre is commercially often known as kevlar®. Both CFRP and AFRP contain 60% in volume of the respective fibre. The mechanical properties of these materials are summarized in Table II.

The last column (i.e.  $E/\rho$ ) has been calculated from the second and the fourth to represent the rigidity per unit of mass of one centimetre cubic of material. This ratio highlights that CFRP has a much higher strength-to-weight ratio than the other two materials considered.

## IV. RESULTS AND DISCUSSION

The numerical simulations performed in this study have been conducted using the algorithms implemented in a commercially available piece of software for modal frequency analysis with finite element method (FEM). A detailed theoretical background for such algorithms can be found for instance in chapter one, two, four and twelve of [13]. In the formulation of the numerical problem, the displacement field has been constrained to be zero on the inner cylindrical surface of the hovering MAV vertical duct. Such a constraint constitutes the boundary condition of the problem.

TABLE II. MECHANICAL PROPERTIES OF MATERIALS

Materials	Mechanical properties			
	Young's modulus, $E/\text{GPa}$	Poisson Ratio, $\nu$	Density, $\rho/\text{g}\cdot\text{cm}^{-3}$	$E\cdot\rho^{-1}/\text{GPa}\cdot\text{cm}^3\cdot\text{g}^{-1}$
Aluminum 2024 <sup>a</sup>	72.4	0.33	2.77	26.14
CFRP <sup>a</sup>	220	0.25	1.70	129.41
AFRP <sup>a</sup>	76	0.34	1.4	54.29

a. Data from [12].

The four geometrical factors examined ( $H$ ,  $R$ ,  $r$ ,  $t$ ) have two levels each, whereas the factor material has three levels. The full factorial design of the numerical experiments therefore encompasses 48 different experimental cases (i.e.  $3 \times 2^4$ ). The first natural frequency numerically estimated as described above has been selected as the response variable. Running a full factorial design has enabled the authors to provide a graphical estimation of both the effects of the factors on the numerically estimated first natural frequency and the second order interaction effects between the factors.

#### A. Main effects

For the main effects, notched box plots have been used to display the results and to estimate graphically if some significant difference between them is present. In a box plots with notches three lines are represented: the 25% percentile, the median and the 75% percentile. The vertical v-groove on the box with apex on the median line is calculated in such a way that if the v-grooves of two boxes are vertically overlapping, and then there is little evidence that the median of the two boxes are significantly different. Also, the distance between the 25% and 75% percentiles allows the investigator to draw conclusions about the dispersion of the response variable for that specific level of the factor under examination. Values of the frequency that lie outside the vertical segments stemming from the box (aka whiskers) should in general deserve further examination in most of the cases. This is to identify possible anomalies occurred in the experimental procedure. More details about box plots and box plots with notches can be found in [14] and [15], respectively. All the figures in this section have been produced using the GPL'd software R, available free of charge at the R-project website (i.e. [www.r-project.org](http://www.r-project.org)). In this qualitative approach, the median has been implicitly used to estimate the centrality parameter of the distribution of values of the simulated frequency for each level of the factor under investigation. In most cases a mean is instead used. However, using the median as estimator of centrality constitutes an approach more robust to the presence of extreme values in the data. As it is shown in the analysis here below, that is the case with the data available in this investigation.

From Fig. 3, the effect of the material on the first natural frequency can be qualitatively estimated. The dispersion of the first natural frequency for each of the materials investigated appears comparable (i.e. comparable vertical heights of the boxes). The CFRP seems to provide a significant higher first natural frequency than Aluminum 2024 (the vertical v-grooves of the correspondent boxes do not overlap). However, a significant difference between Aluminum and AFRP does not appear. Also, it is quite difficult to ascertain whether the difference between AFRP and CFRP is significant: the v-grooves of the correspondent boxes are border-line overlapping. In Fig. 3, six extreme values that depart from the majority of the data for a given material are observed (they are represented as points in Fig.4). A closer scrutiny of these six values reveals that they are obtained in identical combinations of the levels of three factors different from material. Namely, they are obtained

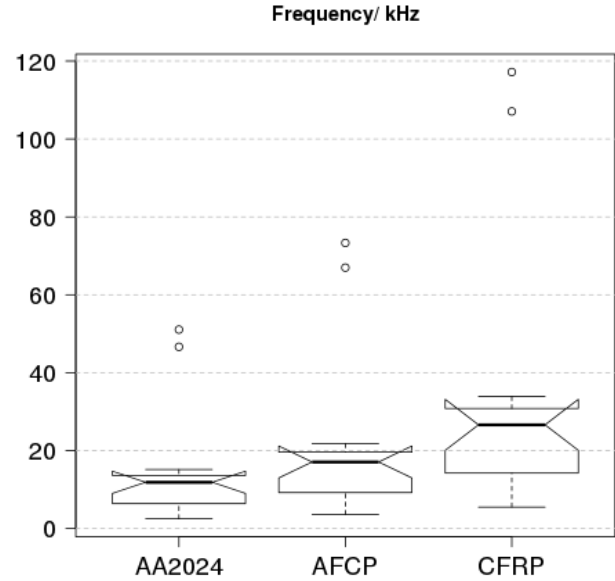


Figure 3. Box plots of the frequency grouped by material.

for a large external radius ( $R=100$  mm), a great thickness of the shell ( $t=10$  mm), a short height of the MAV ( $H=95$  mm). Also it is noticed that for such combination of levels, increasing the internal radius  $r$  from 30 to 60 mm always results in an increment of the simulated first natural frequency for any of the materials considered.

The effect of the external radius on the response variable can be qualitative estimated from Fig. 4. The simulated frequency displays a distribution of values that is more dispersed when using a 100mm external radius than when using 60mm external radius. It then seems more useful to choose a larger external radius, if the design intent is to increase the first natural frequency of the aircraft structure. Also, in Fig. 4, when  $R=100$  mm, two values appears as extreme. A further examination reveals that such a pair of values is among the six outlying points already discussed in Fig.3.

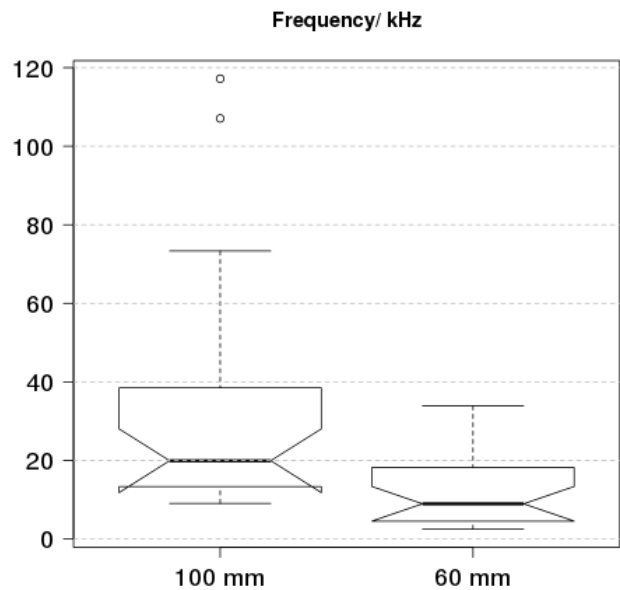


Figure 4. Box plots of the frequency grouped by external radius ( $R$ ).

As displayed in Fig. 5, a very similar situation happens when considering the effect of the height factor on the simulated first frequency. Lower heights of the hovering MAV (i.e. lower  $H$ ) are likely to increase significantly the first natural frequency of the aircraft.

A close examination of the box plots for the internal radius and the thickness did not reveal any significant effect of these two factors on the simulated first natural frequency. For this reason they have not been included in this section. The sole noteworthy comment about them is that the variability of the simulated frequency appears larger at 10 mm thickness than it is at 5 mm.

### B. Second order interaction effects

The effect that a factor, say factor A, have on the simulated first natural frequency may be different when different levels of the other factors involved in the analysis are considered. An interaction plot shows the mean (alias average) of the response for the levels of the factor investigated appearing on the abscissa when the levels of a second factor is also used to subset the data. Joining the means at the same level of the second factor is therefore a means to estimate qualitatively if the two factors in the plot may interact when affecting the response variable. In particular, if the lines so obtained are crossing or tend to cross, then it can be argued that there may be a different effect of the factor on the abscissa on the response variable at different levels of the other factor displayed in the interaction plot. Shortly, there is a strong suspicion of an interaction between the two factors.

As shown in Fig. 6, when  $t=10$  mm, changing the external radius from 60 mm to 100 mm will give a large frequency increment (in the order of 35 kHz). But when  $t=5$  mm, the frequency will have a relatively small increment (in the order of 7 kHz). From a designer perspective, The implications of this finding is that changing the external radius of the outer shell may have quite an unpredictable outcome on the frequency depending on the thickness of the shell itself.

Similar situation appears also for the interaction

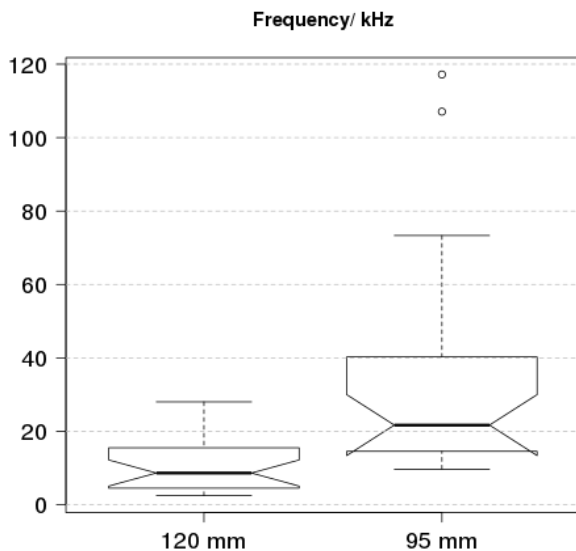


Figure 5. Box plots of the frequency grouped by height ( $H$ ).

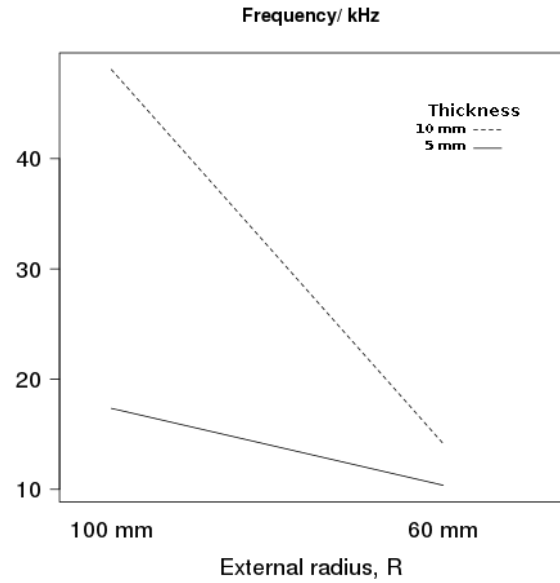


Figure 6. Interaction plot of the external radius ( $R$ ) and the thickness ( $t$ ).

between height and thickness. In Fig. 7, with  $t=10$  mm, when decreasing the height of the MAV from 120 mm to 95 mm, there is a suspicion that the natural frequency increases significantly more than with  $t=5$  mm. From a practical perspective, this finding may for example be useful to a designer that is constrained to use an outer shell with  $t=10$  mm. He/she would know that he/she would probably have better chances of increasing the first natural frequency of the MAV by reducing the height of the structure. But not so, if he/she had been given the constraint  $t=5$  mm.

Fig. 8 can give a qualitative idea of the interaction effect of the internal radius and the thickness. In Fig. 8, when  $t=10$  mm, changing the internal radius from 30 mm to 60 mm leads to an increment of the first natural frequency. But when  $t=5$  mm, the same change of the internal radius causes the first natural frequency to drop. A mild interaction may therefore be present between these

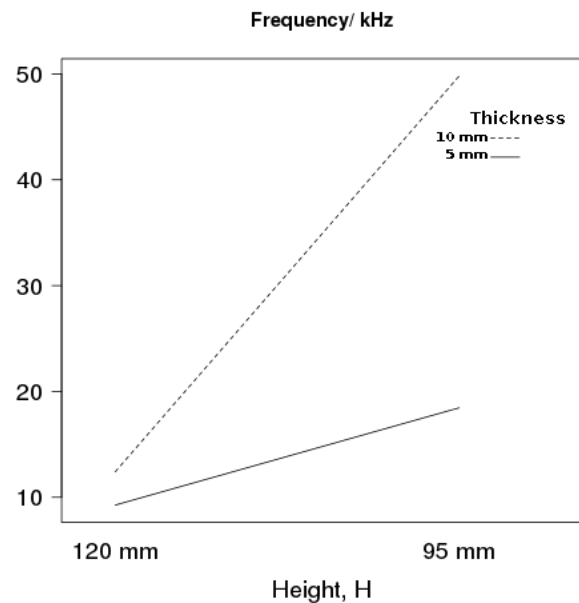


Figure 7. Interaction plot of the height ( $H$ ) and the thickness ( $t$ ).

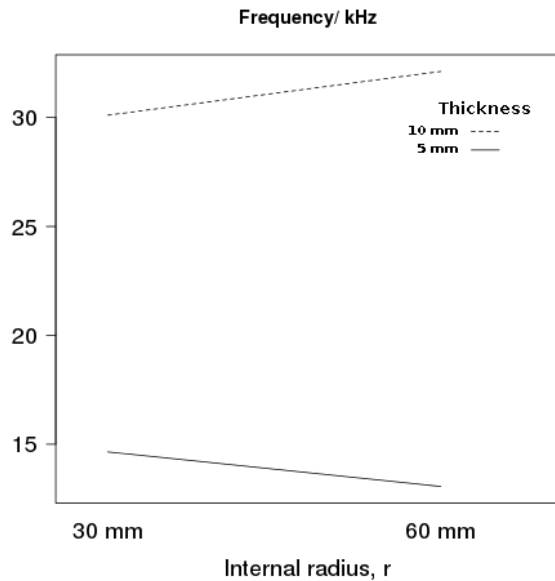


Figure 8. Interaction plot of the internal radius ( $r$ ) and the thickness ( $t$ ).

two factors. All the remaining 10-3=7 second order interaction plots have been analyzed like here above. Yet no interesting additional comments could be elicited from them. For these reason they have been omitted from this section.

## V. CONCLUSIONS

The identification of the natural frequencies of a mechanical structure is an integral part of its design phase. In particular, this should hold also for aerospace applications such as the design of a hovering micro air vehicle. The effect of selected materials and geometrical parameters on the simulated first natural frequency has been analyzed in this study. The numerically simulated values of the first natural frequency were obtained using the modal frequency analysis algorithms implemented into commercially available FEM software. Aluminum alloy 2024, carbon fibre reinforced polymer and aramidic fibre reinforced polymer were selected as materials, whereas four geometry parameters have been changed at two different levels. This setting resulted in 48 cases of a full factorial numerical experiment. The results were analyzed using graphical representations typical of exploratory data analysis (EDA). The main findings of this investigation are:

- Carbon fibre reinforced polymer appeared to provide first natural frequencies higher than the aluminum alloy. Yet no significant difference appeared between the first natural frequency of aluminum alloy and aramidic fibre reinforced polymer. It is unclear however if carbon fibre provides higher first natural frequencies than aramidic fibre when used in the investigated composites.
- If a material has been selected, then the geometry that appears to give a considerable higher first natural frequency is characterized by large radius of the outer shell, great

thickness of such a shell and small height of the MAV.

- Both the radius of the cylindrical duct and the thickness of the shell do not appear to have significant effect on the first natural frequency. Instead, increasing the radius of the outer shell seems to significantly yield higher first natural frequencies. Likewise does reducing the height of the MAV.
- Second order interactions between the investigated factors have been discussed. But not clear interaction effect is apparent.

## REFERENCES

- [1] T.T.H. Ng and G.S.B. Leng, "Application of genetic algorithms to conceptual design of a micro-air vehicle," *Engineering Applications of Artificial Intelligence*, vol. 15, pp. 439-445, September 2002.
- [2] Bor-Jang Tsai and Yu-Chun Fu, "Design and aerodynamic analysis of a flapping-wing micro aerial vehicle," *Aerospace Science and Technology*, vol. 13, pp. 383-392, October-November 2009.
- [3] Takeo Kanade, Omead Amidi, and Qifa Ke, "Real-time and 3D vision for autonomous small and micro air vehicles," 43rd IEEE Conference on Decision and Control Bahamas, vol. 2, pp. 1655-1662, December 2004.
- [4] Richard J. Bachmann, Frank J. Boria, Ravi Vaidyanathan, Peter G. Ifju, and Roger D. Quinn, "A biologically inspired micro-vehicle capable of aerial and terrestrial locomotion," *Mechanism and Machine Theory*, vol. 44, pp. 513-526, March 2009.
- [5] Jean Michel Pflimlin, Philippe Soueres, and Tarek Hamel, "Hovering flight stabilization in wind gusts for ducted fan UAV," 43rd IEEE Conference on Decision and Control Bahamas, vol. 4, pp. 3491-3496, December 2004.
- [6] H. Rajabi, M. Moghadami, and A. Darvizeh, "Investigation of microstructure, natural frequencies and vibration modes of dragonfly wing," *Journal of Bionic Engineering*, vol. 8, pp. 165-173, June 2011.
- [7] D.N. Paliwal, H. Kanagasabapathy, and K.M. Gupta, "Vibrations of an orthotropic shallow spherical shell on a Kerr foundation," *International Journal of Pressure Vessels and Piping*, vol. 64, pp. 17-24, 1995.
- [8] C. Wang and J.C.S. Lai, "Prediction of natural frequencies of finite length circular cylindrical shells," *Applied Acoustics*, vol. 59, pp. 385-400, April 2000.
- [9] Eric N. Johnson and Michael A. Turbe, "Modeling, control, and flight testing of a small ducted fan aircraft," *AIAA Guidance, Navigation, and Control Conference and Exhibit California*, pp. 1-23, August 2005.
- [10] <http://www.aviationweek.com/aw/>.
- [11] [http://www.aviationmetals.net/2024\\_aluminum.php](http://www.aviationmetals.net/2024_aluminum.php).
- [12] William D. Callister, Jr, *Materials Science and Engineering: An Introduction*, 6th ed., New York; Chichester: Wiley, 2003, pp. 737-745.
- [13] Robert D. Cook, *Concepts and Applications of Finite Element Analysis: A Treatment of the Finite Element Method as used for the Analysis of Displacement, Strain, and Stress*, New York; London (etc.): Wiley, 1974.
- [14] John W. Tukey, *Exploratory data analysis*, Addison-Wesley, 1977, section 2c.
- [15] Robert McGill, John W. Tukey, and Wayne A. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, pp. 12-16, February 1978.

# Finite Element Investigation of Nano-indentation of coated Stainless Steel

Qiang Xu, Chulin Jiang, Dezheng Liu, Yongxin Pang, Simon Hodgson

*School of Science and Engineering*  
Teesside University  
Middlesborough, Tees Valley, TS1 3BA  
q.xu@tees.ac.uk

**Abstract**—An finite element analysis was used to investigate the nano-indentation testing process in different kinds of coating on 316 Stainless Steel, in order to determine the properties of coating material and to investigate the influence of different of the material coating materials on 316 Stainless Steel. The finite element analysis was based elastic-plasticity properties of material. The results were three main points: 1) the simulation force-depth relationship was agreed with the experimental force-depth relationship for 316 stainless steel; 2) the range of Modulus of Elasticity for specific coating was determined as between 5GPa to 60GPa; 3) the influence of the thickness of coating material on the force indentation depth relationship was investigated. The project offers suggestions to further design of coating process research.

**Keywords:** coating material and property, nano-indentation, finite element analysis

## I. INTRODUCTION

The nano-indentation is an experiment for testing properties for coating and film and finite element method has been used in such research, for example [1]. Part of the objectives of the current project [2, 3] was specified as:

1) Testing for plate with only 316 Stainless Steel was simulated with FE, comparing with the result of experiment in nano-indentation in order to validate the correctness and accuracy of the FE analysis;

2) Determination the Modulus of Elasticity for specific ceramics coated 316 stainless steel, comparing with the result of experiment in nano-indentation;

3) Investigation the influence coating thickness on the force-indentation depth.

This paper reported the research work in order achieving the above objectives including the background information, the FE model developed and a series of case study.

## II. BACKGROUND INFORMATION

**Stainless Steel:** The property of stainless steel at room temperature is shown in Table 1 and Figure 1.

Table 1 Basic Property of Stainless Steel [4]

Modulus of Elasticity	Poisson Ratio	Yield Stress
193GPa	0.263	300MPa

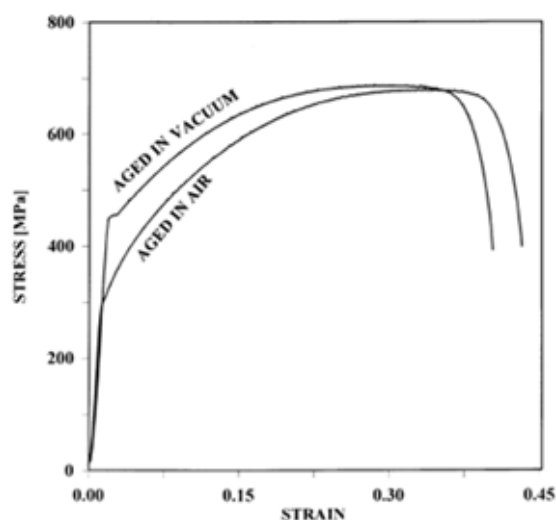


Figure 1: Relationship between stress and strain [5]

**Indenter:** The indenter was diamond. The geometry of the indenter was shown in Figure 1: Geometry of Indenter; the properties of the diamond was shown by Table 2: Synthetic diamonds values.

Table 2. Synthetic Diamonds Property

Young's Modules	Poisson's Ratio
1220 GPa	0.2

It is understood that the ceramic material is of brittle fracture nature and it is idealized as rigid plasticity for simplification and due to the limit of the I-DEAS software [6]. The base material stainless steel is modeled as elastic-plasticity accurately. It is also assumed no friction between the coating and indenter tip.



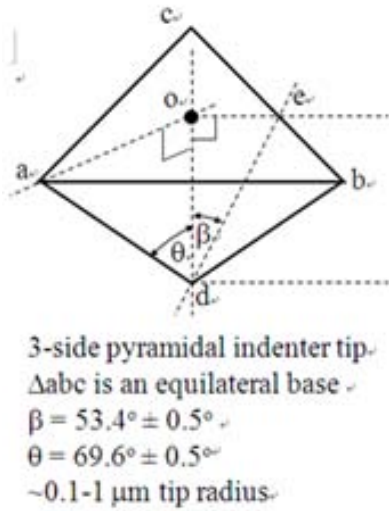


Figure 2. The shape and dimension of 3-side pyramidal indenter tip

The FE analyses were organized as:

1) *Validating FE Model*

Experimental nano-indentation test data of 316 stainless steel obtained from an industrial funded research project at Teesside University was used for validating the FE model developed.

2) *Determination of the value of Modulus of Elasticity for the coating material*

A series of trial and error FE analysis were conducted with varying Young's modules of coating material. The FE obtained displacement and force relationship were used to determine the range of the Modulus of Elasticity through comparison with experimental data.

3) *Influence of coating thickness*

A series of FE analysis with varying thickness of coating was conducted in order to obtain the quantitative results of its influence on the force-indentation depth.

### III. RESULTS

#### 3.1 Validating FE Model

The main aim was to validate the method of finite element techniques for the investigation of the nano-indentation test through the comparison of the predicted and the experimental force-depth relationship of 316 Stainless Steel. The FE model developed is illustrated in Figure 3.

The solution of model was illustrated as Fig. 3: the Model of Nano-indentation for 316 Stainless Steel. From this Figure, it was known that the maximum stress was at the top and it is contact stress, which was brought by the indenter.

Through the comparison of the FE prediction and experimental measurement shown in Figure 4, it is clearly that the FE prediction agrees very well with experimental data which confirms that the correctness and the accuracy of the FE model developed.

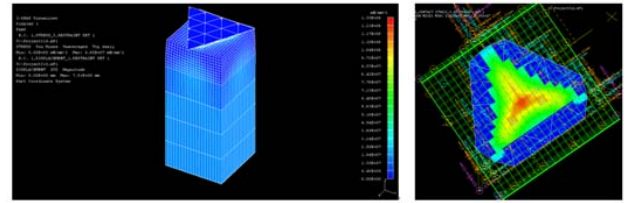


Figure 3 The FE Model and stress distribution

Table 3: The FE prediction of force and displacement during indenting pure stainless steel

Depth	Force
0nm	0mN
150nm	3.624mN
300nm	13.64mN
450nm	27.8mN
600nm	44.8mN
750nm	64.0mN

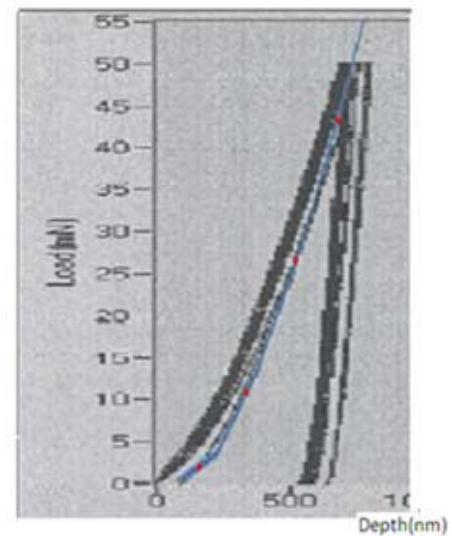


Figure 4 Comparison of FE predicted and experimental measured force-indentation depth relationship

#### 3.2 Determination of the Modulus of Elasticity for specific coated 316 stainless steel

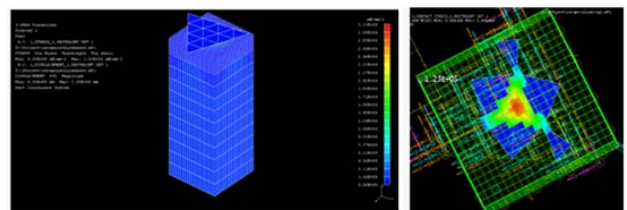


Figure 5 The FE Model and Contact of Top surface



Table 4 Trial values for the lower and upper limit of E

Lower Limit E	4.5GPa	5GPa	5.5GPa
Upper limit E	55GPa	60GPa	62GPa

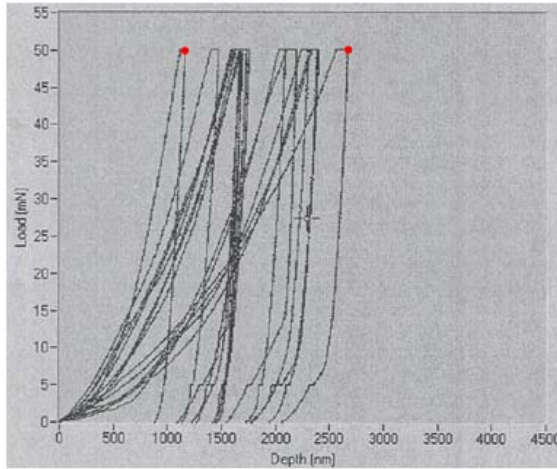


Figure 6 The load vs indentation depth with Modulus of Elasticity of 5 GPa and 60 GPa

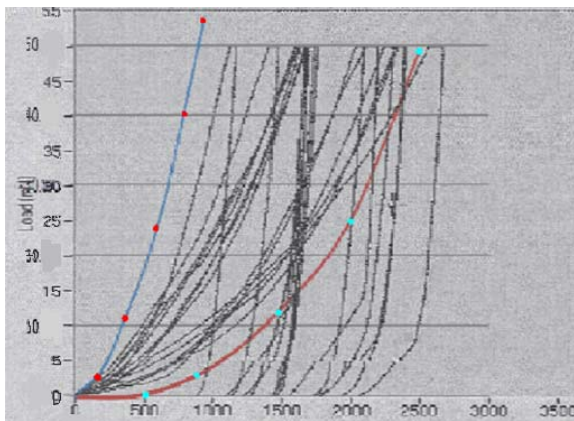


Figure 7 The force and indentation depth relationship

The FE model developed for this investigation and the top contact area are shown in Fig. 5. Fig. 6 shown the indentation depths for a given load assuming the coating material with E of 5 GPa and 60 GPa, which corresponded to the upper and lower limit of experimental observation. Fig. 7 shown the force indentation depth relationship during the whole indentation process, which clearly demonstrated a very good agreement between the FE prediction and experimental observation.

### 3.3 Influence of coating thickness

Three coating thickness were chosen, namely, 4  $\mu\text{m}$ , 5  $\mu\text{m}$ , and 6  $\mu\text{m}$  and the two values of E (5 GPa and 60 GPa) were used in the FE analysis. The typical FE Model is shown in Fig. 8 and the results are shown in Fig 9.

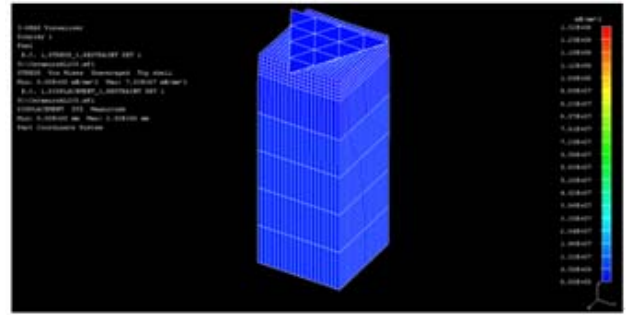


Figure 8 Typical FE Model

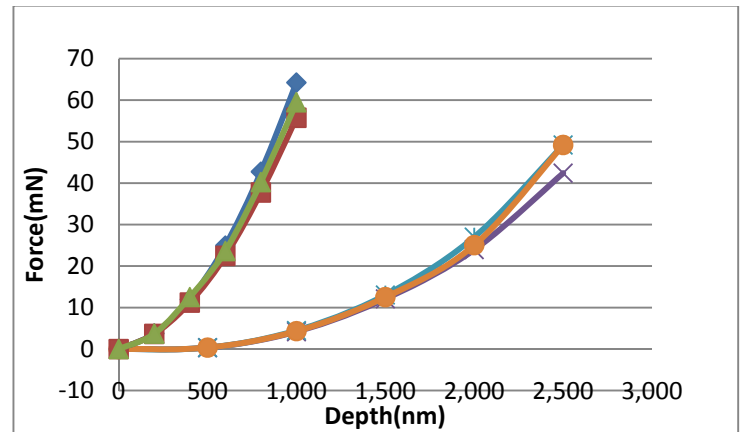


Figure 9 The effect of thickness on the force-displacement relationship:  $\diamond$ ,  $\circ$ , and  $\square$  for 4  $\mu\text{m}$ , 5  $\mu\text{m}$ , and 6  $\mu\text{m}$  respectively

It is interesting to note from Fig. 9 that the thickness seems no significant influence on the force-indentation depth. This is supported by the fact observed from the Fig. 10, Fig. 11, Fig. 13 that the deformation under the tip of indenter is very localized within the coating, further increase of its thickness would not affect the relationship.

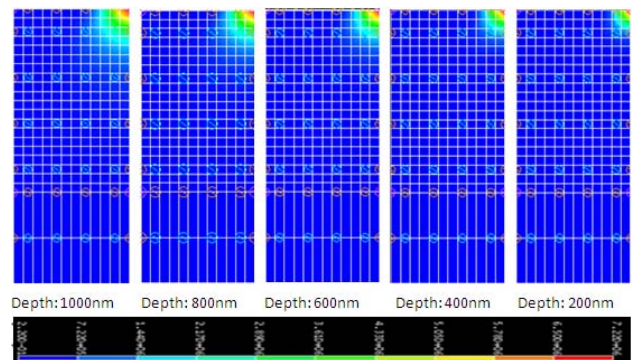


Figure 10a Effective Strain Distribution for Coating with 4 $\mu\text{m}$  Thickness and 5GPa Modulus of Elasticity

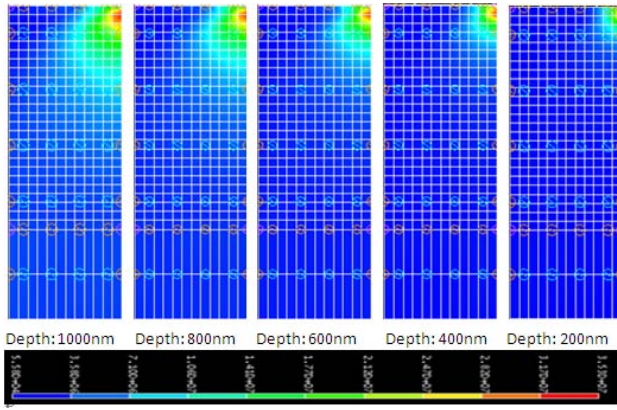


Figure 10b Effective Stress Distribution for Coating with 4 $\mu$ m Thickness and 5GPa Modulus of Elasticity

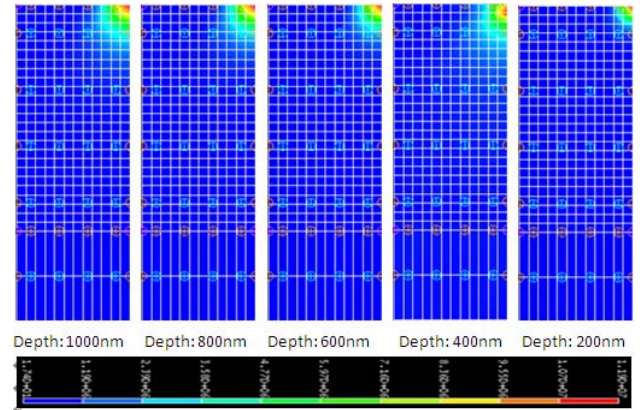


Figure 12a Effective Strain Distribution for Coating with 6 $\mu$ m Thickness and 5GPa Modulus of Elasticity

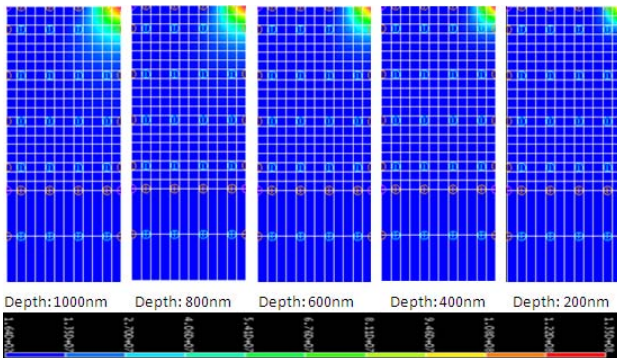


Figure 11a Effective Strain Distribution for Coating with 5 $\mu$ m Thickness and 5GPa Modulus of Elasticity

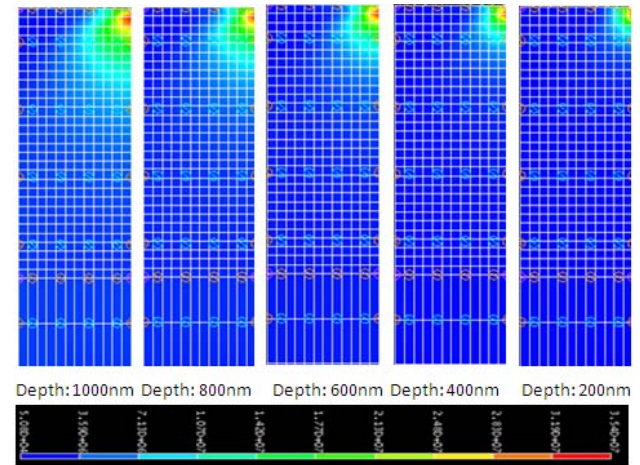


Figure 12b Effective Stress Distribution for Coating with 6 $\mu$ m Thickness and 5GPa Modulus of Elasticity

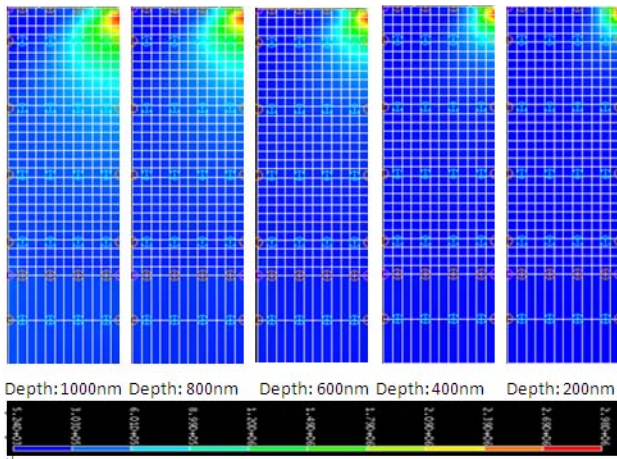


Figure 11b Effective Stress Distribution for Coating with 5 $\mu$ m Thickness and 5GPa Modulus of Elasticity

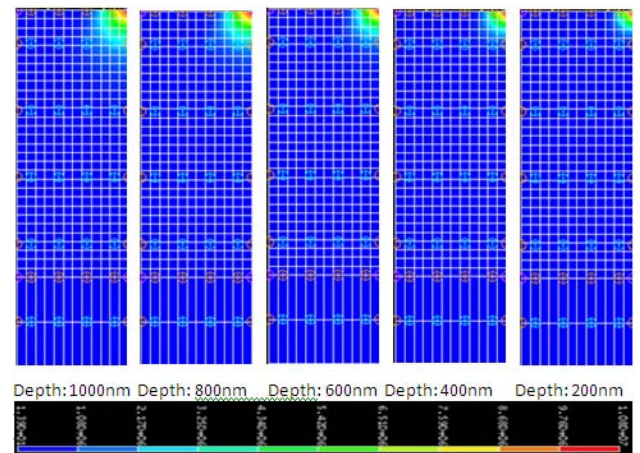


Figure 13a Effective Strain Distribution for Coating with 6 $\mu$ m Thickness and 5GPa Modulus of Elasticity



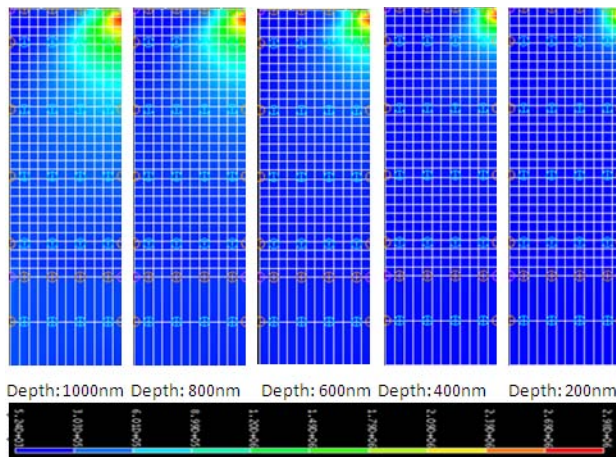


Figure 13b Effective Stress Distribution for Coating with 6 $\mu$ m Thickness and 60GPa Modulus of Elasticity

#### IV. CONCLUSION

The overall conclusions are:

- 1) The FE analysis based on contact and elastic-plasticity model is reasonably accurate to depict the force-indentation depth;
- 2) The use of such FE model has been successfully used to determine the lower and upper limit of E for coating material;
- 3) It was revealed that when the coating thickness is in the order of 4 to 6  $\mu$ m, the thickness of coating has not significant influence on the force-indentation relationship. It is suggested that there is a lower bound for coating thickness which has practical significance in developing coated functional material.

#### V. REFERENCES

- [1] J. M. Lee, et al, "Identification of the bulk behaviour of coatings by nano-indentation test and FE-analysis and its application to forming analysis of the coated steel sheet," Journal of Materials Processing Technology, Vol. 187-188, pp.309-313, 2007
- [2] C. Jiang, Finite Element Investigation of Nano-indentation of coated stainless steel, final year project thesis, School of Science and Engineering, Teesside University, 2011
- [3] D. Liu, Finite Element Investigation of Nano-indentation of coated stainless steel, final year project thesis, School of Science and Engineering, Teesside University, 2011
- [4] A. Abduluyahed, and K. J. Kurzydłowski, "Tensile properties of a type 316 stainless steel strained in air and vacuum," Materials Science and Engineering A, vol.256, issue 1, pp.34-38, 1998
- [5] Micro Materials Ltd, [www.micromaterials.co.uk](http://www.micromaterials.co.uk)
- [6] I-DEAS

# The Development and Validation of Multi-axial Creep Damage Constitutive Equations for P91

Qiang Xu, Mark Wright, Qihua Xu

*School of Science and Engineering*

Teesside University

Middlesborough, Tees Valley, TS1 3BA

q.xu@tees.ac.uk

**Abstract**—The development and validation of multi-axial creep damage constitutive equations for A369 FP91 steel at 625°C was presented. Three aspects of research work were reported: 1) the development of a new set of creep damage constitutive equations based on Xu's formulation [1], 2) the validation on this newly developed and an existing KRH type of constitutive equations were validated under plane stress and plane strain conditions, and 3) discussion and conclusion.

**Keywords:** creep damage; constitutive equations; validation; multi-axial formulation

## I. INTRODUCTION

Creep design of metallic components operating at high temperature is largely carried out using design code/methods (e.g. BS 5500, and ASME Subsection NH). Flexibility is provided with some codes (e.g. BS 5500) for the designer to use by analysis methods and it has been explored with creep continuum damage mechanics together with the finite element method.

The phenomenological approach of creep damage mechanics can be broadly classified into weak coupling and strong coupling between damage and deformation. In the case of weak coupling the effect of material damage in elastic properties is disregarded and a coupling is established by introducing the damage variables into the constitutive equations with the concept of effective states variables. Within the weak coupling approach, a set of creep damage constitutive equations for uni-axial tension is generalised for multi-axial applications. The success of this approach depends on the development of a set of appropriate constitutive equations capable of depicting the observed multi-axial material behaviour, which will be assessed through validation. That is the concern of this paper.

Specifically to the formulation and validation methodology, Xu proposed a new formulation for multi-axial creep damage constitutive equations [1, 2] and also proposed an improved validation methodology [2] and reported its application to 0.5Cr0.5Mo0.25V ferritic steel [2].

It was noted very limited work published on the review of constitutive equations except a recent publication [3]. It is evident that there is a degree of confusion between calibration and validation among research community.

The P91 as one of 9Cr martensitic steels was developed as a result of the demand for ferritic steels with higher creep strength, but to-date there has been a somewhat limited in-plant experience of their long-term performance, especially with thick section components such as headers and steam pipework [4].

According to [5] it was reported that reduction of creep rupture strength on high Cr steels was investigated: 1) creep damage in the weld joint preferentially accumulated in fine grain region in heat affected zone (HAZ) where creep deformation resistance is lower than other portions [6]; 2) creep void initiation and growth are accelerated due to multi-axial stress states in the HAZ [7], 3) it was also found [8] that creep rupture life of weld joints failed at the HAZ is approximately 1/5 of the base metal and creep strain concentrated in the HAZ of the weld joint based on a finite element creep analysis using the three materials weld joint specimen model consisting of a base metal, a weld metal and a HAZ. Therefore so-called Type IV cracking occurs in actual components.

Recently, research attempted to use computational creep damage mechanics to investigate problems related to P91 pipes and weldments where different type of creep damage constitutive equations were used [9, 10], however, there was not adequate consideration of validation and probably a degree of confusion about calibration and validation.

This paper reports a review of creep deformation and damage, formulation, validation methodology and practice, the recent progress on the development of new set of creep damage constitutive equations, validation for the specific material.

This paper not only demonstrates the capability of this new set of constitutive equations, but also reveals the deficiency in the KRH type of constitutive equations; this discovery is similar to the finding of previous work on 0.5Cr0.5Mo0.25V ferritic steel [2]. Thus, the author addresses again that in the developing damage mechanics field, it is important and necessary to distil the information and conclusions cautiously prior to developing, accepting, and applying any theory and constitutive equation, either new or old (including this one).

## II. CREEP DEFORMATION AND DAMAGE [1]

The creep deformation is typically divided into primary, second, and tertiary creep; while the damage process is often understood (with great simplification) to

be a process of nucleation, cavity growth and coalescence. In martensite/ferritic steels damage is often associated with substructure coarsening. Uni-axial constitutive equations capable of describing primary, second and tertiary creep have been developed. If only one damage variable is chosen, the creep strain rate  $\dot{\epsilon}$  and damage rate  $\dot{\omega}$  by

$$\dot{\epsilon} = f(\sigma, \omega) \text{ and } \dot{\omega} = g(\sigma, \omega) \quad (1)$$

By appropriate selection of the function  $f$  and  $g$ , as well as the critical value of damage, it is possible to represent the tertiary creep and to produce a stress-lifetime relationship consistent with experimental observations. One of the examples is

$$\dot{\epsilon} = G \left[ \frac{\sigma}{1-\omega} \right]^n \text{ and } \dot{\omega} = C \frac{\sigma^\chi}{(1-\omega)^\phi} \quad (2)$$

where  $G$ ,  $C$ ,  $n$ ,  $\chi$  and  $\phi$  are material constants. The effective stress concept is used and the material is deemed to fail when the value of the damage variable reaches its critical value of 1.

Then primary creep could be included by extra hardening function  $H$ , such as being proposed and used. For example, the constitutive equations form for uni-axial condition is given as:

$$\dot{\epsilon} = A \sinh \left( \frac{B\sigma(1-H)}{(1-\phi)(1-\omega)} \right) \quad (1)$$

$$\dot{H} = \frac{h}{\sigma} \left( 1 - \frac{H}{H^*} \right) \dot{\epsilon} \quad (4)$$

$$\dot{\phi} = \frac{K_c}{3} (1-\phi)^4 \quad (5)$$

$$\dot{\omega} = C \dot{\epsilon} \quad (6)$$

where  $A$ ,  $B$ ,  $h$ ,  $H^*$ , and  $D$  are material constants.

#### A. KRH multi-axial Formulation

The KRH type multi-axial creep damage constitutive equations are given as [10]:

$$\dot{\epsilon}_e = A \sinh \left( \frac{B\sigma_e(1-H)}{(1-\phi)(1-\omega)} \right) \quad (7)$$

$$\dot{H} = \frac{h}{\sigma} \left( 1 - \frac{H}{H^*} \right) \dot{\epsilon}_e \quad (8)$$

$$\dot{\phi} = \frac{K_c}{3} (1-\phi)^4 \quad (9)$$

$$\dot{\omega} = C \dot{\epsilon}_e \left( \frac{\sigma_1}{\sigma_e} \right)^\nu \quad (10)$$

where  $\nu$  is the stress state index.

#### B. New Multi-Axial Formulation

The approach originally proposed by Xu [1, 2] was adopted here and the multi-axial creep damage constitutive equations are given as:

$$\dot{\epsilon}_e = A \sinh \left( \frac{B\sigma_e(1-H)}{(1-\phi)(1-\omega_d)} \right) \quad (11)$$

$$\dot{H} = \frac{h}{\sigma_e} \left( 1 - \frac{H}{H^*} \right) \dot{\epsilon}_e \quad (12)$$

$$\dot{\phi} = \frac{K_c}{3} (1-\phi)^4 \quad (13)$$

$$\dot{\omega} = C \dot{\epsilon}_e f_2 \quad (14)$$

$$\dot{\omega}_d = \dot{\omega}^* f_1 \quad (15)$$

#### C. Specific Forms

$$f_1 = \left( \frac{2\sigma_e}{3\sigma_1} \right)^a \exp \left( b \left( \frac{3\sigma_m}{\sigma_s} - 1 \right) \right) \quad (16)$$

$$f_2 = \left( \exp \left( p \left( 1 - \frac{\sigma_1}{\sigma_e} \right) + q \left( \frac{1}{2} - \frac{3\sigma_m}{2\sigma_e} \right) \right) \right)^{\frac{1}{2}} \quad (17)$$

### III. VALIDATION METHOD [1]

First, an adequate validation should address (1) what needs to be assessed and (2) under what conditions. With clear understanding of the two fundamental consistency requirements addressed above, it is clear that an adequate validation should be designed and conducted considering:

(1) The items: (a) creep strain rate; and (b) damage evolution;

(2) The stress states: (a) creep curves under uni-axial conditions; (b) multi-axial stress states under proportional loading conditions; and (c) multi-axial states of stress under non-proportional loading conditions.

If a set of creep damage equations is integrated from virgin state to failure, it will produce:

$$\omega = \omega_f, \quad \epsilon = \epsilon_f^{total}, \quad \dot{\epsilon} = \dot{\epsilon}_f^{total} \quad (18)$$

where  $\omega_f$  is the critical value of damage and  $\epsilon_f$  is effective creep strain. These results will be used in validation.

Ideally, the conditions should include proportional and non-proportional loading under multi-axial stress states. Compromise may have to be made due to the constraint imposed by the difficulty to conduct the required experiments and the cost involved, which is not the same as ignorance. Previous practice was not adequate in either the items to be assessed or the range of states of stress.

Practical validation method is proposed as:

(1) To check isochronous rupture loci under plane stress and plane strain states with proportional loading conditions;

(2) To check strain at failure under plane stress and plane strain states with proportional loading conditions;

(3) To check typical creep curves under plane stress and plane strain states with proportional loading conditions;

(4) To check the damage development, creep strain development, strain at failure and lifetime for multi-axial stress states complex (or non-proportional) loading condition. One way to achieve this is notched bar test.

In steps 1 and 2, the plane stress and plane strain stress states are selected to present multi-axial stress states under proportional loading conditions. It is suggested that all the

material constants should be determined in the first two steps. Step 3 intends to further check the coupling of damage and creep strain. Step 4 validates the constitutive equations under multi-axial non-proportional loading conditions. It is clear that previous practice is not adequate as it ignored the need to include strain at failure under plane stress states with proportional loading conditions and did not consider plane stress states with proportional conditions. This paper will present validation results on the first three accounts.

#### IV. RESULT

These two sets of multi-axial creep damage constitutive equations are validated in terms of lifetime, ratios of strain at failure, and creep curves, which corresponds to the first three steps of practical validation method described in above section.

The isochronous rupture loci and ratios of strain at failure for KRH formulation are presented in Figure 1 and Figure 2, respectively, while typical results for the new formulation are presented in Figure 3 and Figure 4. The legends in Figure 1 and Figure 2 are the stress state sensitivity index  $\nu$ , while the legends in Figure 3 and Figure 4 are parameter  $q$ . Typical results of creep curves and damage evolution under plane stress condition (proportional loading) are shown in Figure 5 and Figure 6. A comparison of creep curves under pure shear condition is given in Figure 7.

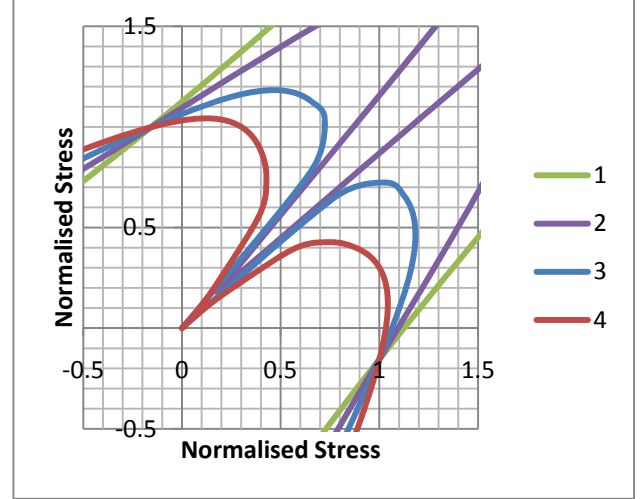
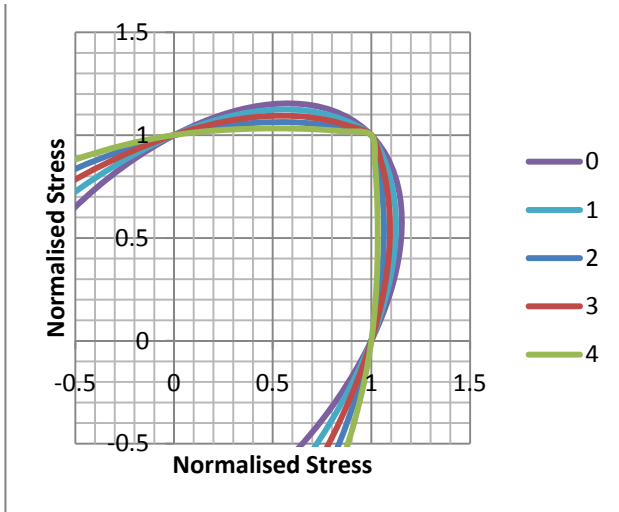


Figure 1. Isochronous rupture loci for KRH formulation (top for plane stress condition and bottom for plane strain condition for  $\nu = 0, 1, 2, 3, \text{ and } 4$ ). The legends are for stress state index  $\nu$ .

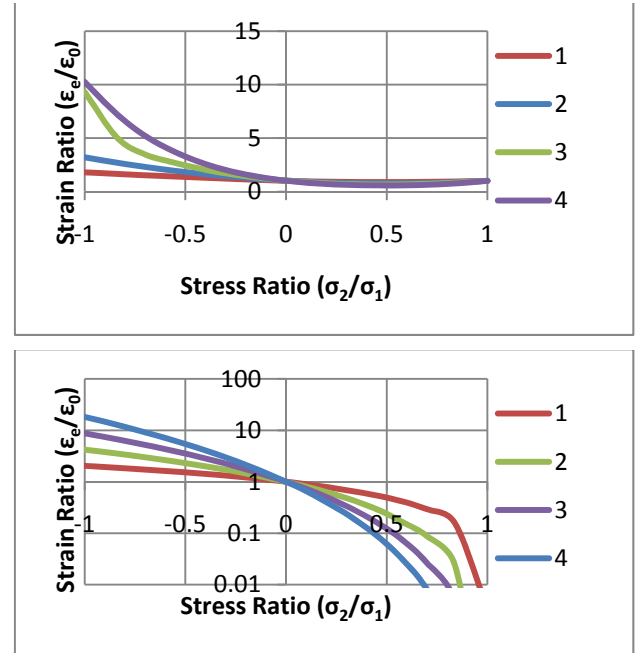


Figure 2. Ratios of strain at failure for KRH formulation (top for plane stress condition and bottom for plane strain condition for  $\nu = 0, 1, 2, 3, \text{ and } 4$ ). The legends are for stress state index  $\nu$ .



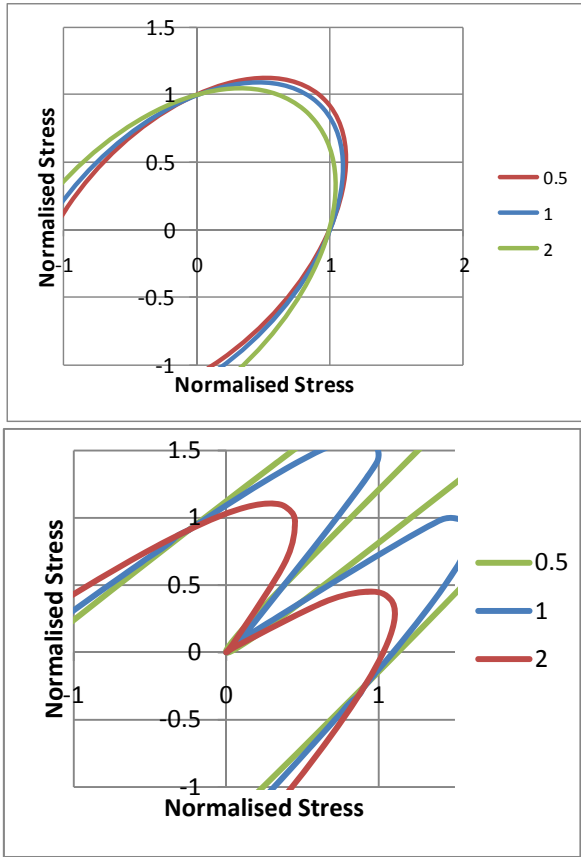


Figure 3. Ratio of strain at failure for new formulation: (a) for plane stress condition and (b) for plane strain condition.  $A = 0$ ,  $b = 0$ ,  $p = 0$ ,  $q = 0.5, 1$  and  $2$ . The legends are for stress parameter  $q$ .

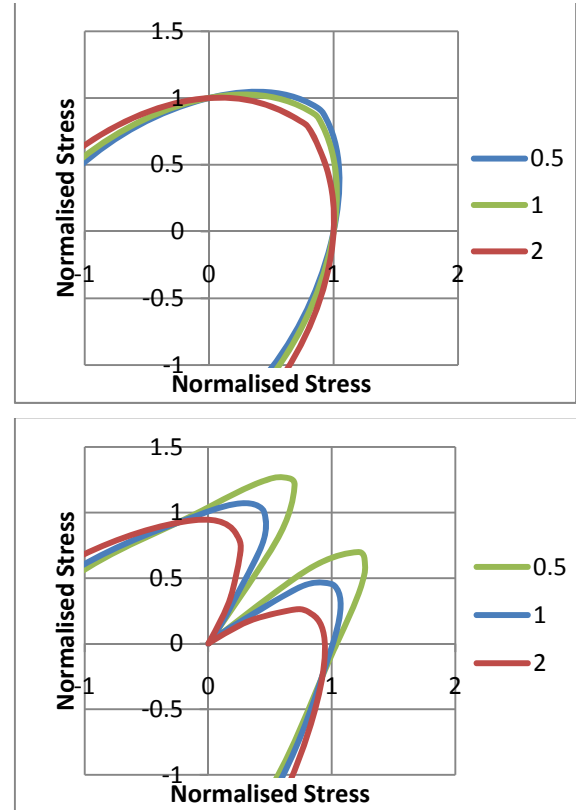


Figure 5. Isochronous rupture loci for new formulation with  $a = 0$ ,  $b = 1$ ,  $p = 1$ ,  $q = 0.5, 1$ , and  $2$ . The legends are for stress parameter  $q$ .

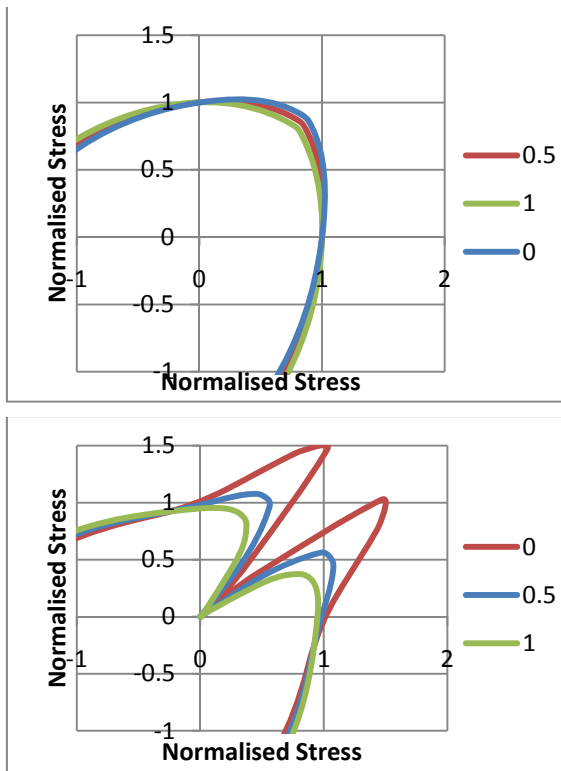
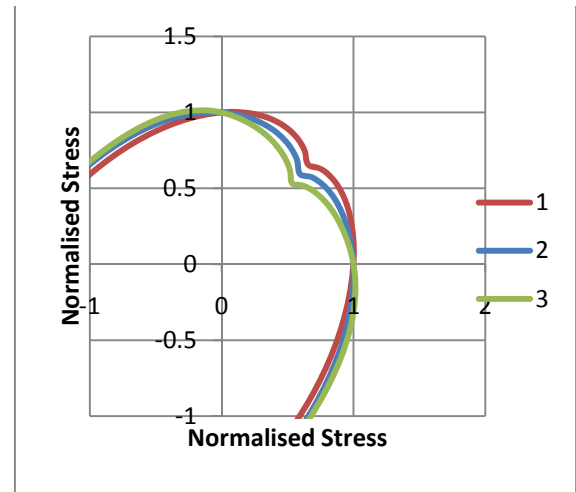


Figure 4. Isochronous rupture loci for new formulation with  $a = 0$ ,  $b = 2$ ,  $p = 1$ ,  $q = 0, 0.5$  and  $1$ . The legends are for stress parameter  $q$ .



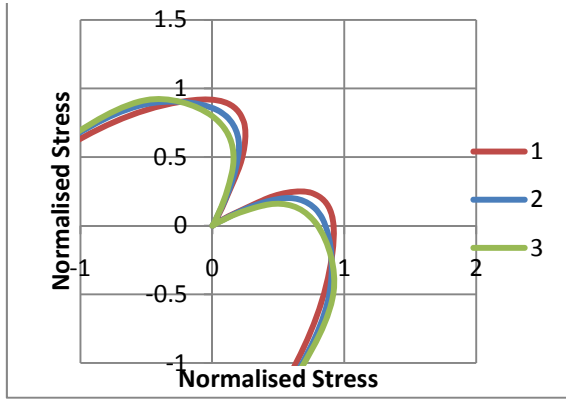


Figure 6. Isochronous rupture loci for new formulation with  $p = 2.5$ ,  $q = 1$ ,  $a = 2$  and  $b = 1, 2$  and  $3$ . The legends are for stress parameter  $b$ .

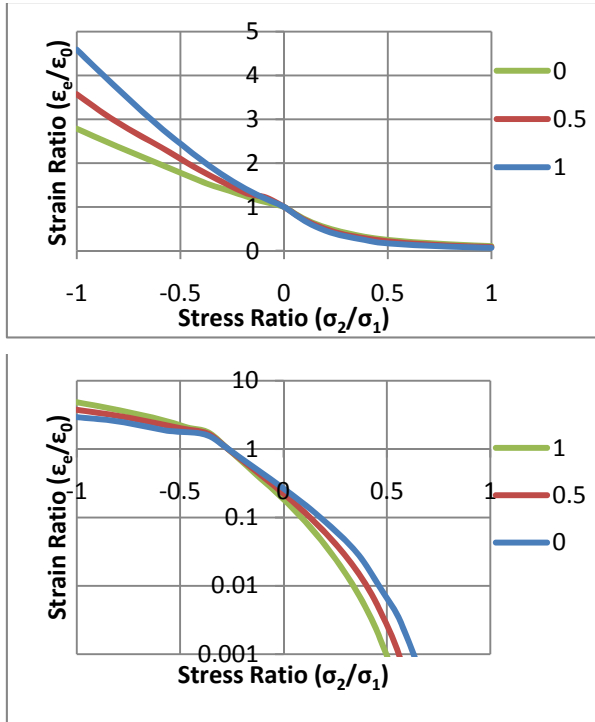


Figure 7. Ratios of strain at failure for new formulation (top for plane stress condition and bottom for plane strain condition with  $q = 0, 0.5$  and  $1$ ). The legends are for stress state parameter  $q$ .

## V. DISCUSSION AND CONCLUSION

### A. KRH Formulation

#### Criterion 1

The plane stress and plain strain isochronous rupture loci (Figure 1) created by using the KRH [11] type of constitutive model and utilised by Hyde et al. [10] demonstrates the inconsistency of the model. Hyde et al. advocate the use of an arbitrary value of one for the stress sensitivity index  $v$  as the claim is made the result is “weakly dependent” upon this value. When the value of  $v$  is one, the plane strain isochronous rupture loci (Figure 2) shows such an increase in creep strength, that this case is simply unrealistic. The plane stress and plane strain isochronous rupture loci demonstrate that the KRH constitutive model cannot simultaneously satisfy the plane stress and plane strain states with a single variable, this is

in line with the early finding on a different material [1]. The plane strain state was not included in [11] and no validation at all is reported [10].

The explicit nature of stress state index  $v$  is described in [11] as “a value that minimises the difference between experimental and computed lifetimes”. The calibration of the stress sensitivity index  $v$  was accomplished by creating a ratio of actual and computed failure times of several notched bar tests and internally pressurised cylinders. The stress state index (within 10%) of where most of these failures occurred was then proposed as the appropriate value. The analysis in this paper demonstrates this number to be more convenient than representative of the multi-axial stress state.

#### Criterion 2

The plane stress and plane strain results were plotted as a ratio of principal stresses and strain at failure (Figure 2 and Figure 3). The results further demonstrate the variance and inconsistency as a result of using the stress state index. The range of the plots is more representative of the stress sensitivity index and its mathematical effect, rather than the relative strain at these points. As the value  $v$  increases, the variance in terms of strain becomes increasingly evident and unrealistic.

#### Criterion 3

The KRH formulation will produce equal life time for uni-axial tension and bi-axial tension which is general inconsistent with experimental observation.

### B. Xu's Formulation

Evaluating the results determined by the Xu formulation using the validation methodology:

#### Criterion 1

A range of plane stress and plane strain isochronous rupture loci are shown in Figs 4 to 5. The optimal results were determined in this analysis using the values proposed by Xu [1] is shown in Figure 6. The plane stress and plane strain isochronous rupture loci (Figure 8 and Figure 9) simultaneously satisfy the plane stress and plane strain states with consistent values of normalised stress in both cases.

#### Criterion 2

Figure 7 plots the range of stress ratios determined by the Xu's formulation. The plots demonstrate a good correlation between the strain at failure across the range of stress ratios and higher accuracy over the range of the analysis.

#### Criterion 3

The biaxial tension lifetime is significantly short than that of uni-axial one.

### C. Summary

Krausz and Krausz summarise their findings in a review of constitutive models and state “there are many different facets of the same problem and as many answers,

the right one is the one that gives the most practical solution, the one that best serves the specific situation” [12]. When compared, the results of the analyses demonstrate a significant increase in the accuracy of the constitutive model and the data produced fully supports this. The isochronous rupture loci reveal the deficiency in RKH formulation and demonstrate the improvement obtained in new formulation proposed by Xu. The criteria of the validation methodology ensured a consistent and fair comparison between the models with demonstrable results.

It is strongly emphasized the need of Experimental data and analysis to satisfy non-proportional loading conditions before the general applicability of a set of constitutive equations is applied for general case study.

#### REFERENCES

- [1] Q. Xu, “The development of validation methodology of multiaxial creep damage constitutive equations and its application to 0.5Cr0.5Mo0.25V ferritic steel at 590°C,” *Nuclear Engineering and Design*, Vol.228, pp.97-106, 2004.
- [2] Q. Xu, “Creep damage constitutive equations for multiaxial states of stress for 0.5Cr0.5Mo0.25V ferritic steel at 590°C,” *Theoretical and Applied Fracture Mechanics*, vol.2, pp.99-107, 2001.
- [3] H. T. Yao, F. Z. Xuan, Z. Wang, S. T. Tu, “A review of creep analysis and design under multiaxial stress states,” *Nuclear Engineering and Design*, vol. 237, pp.1969-1986, 2007.
- [4] A. Shibli, and F. Starr, “Some aspects of plant and research experience in the use of high strength martensitic steel P91,” *International Journal of Pressure Vessels and Piping*, vo.84, pp.114-122, 2007.
- [5] T. Ogata, T.Sakai, and M. Yaguchi, “Damage characterisation of a P91 steel weldment under uniaxial and multiaxial creep,” *Materials Science and Engineering: A*, vol.510-511, pp.238-243, June 2009.
- [6] T. Ogata, and M. Yaguchi, *Proceedings of ECCC Creep Conference on Creep & Fracture in High Temperature Components* London, UK, September 12–14., DEStech Publications, Inc., Lancaster, Pennsylvania, USA (2005), pp. 909–917.
- [7] M. Tabuchi, T. Watanabe, K. Kubo, M. Matsui, J. Kinugawa, and F. Abe, *J. Soc. Mater. Sci.* 50 (in Japanese), 2001
- [8] T. Ogata, T. Sakai, M Yaguchi, *Proceedings of the Third International Conference on Integrity of High Temperature Welds* London, UK, April 24–26, IOM Communication Ltd. (2007), pp. 285–294.
- [9] T. H. Hyde, A. A. Becker, W. Sun, and J.A. Williams, “Finite-element creep damage analyses of P91 pipes,” *International Journal of Pressure Vessels and Piping*, vol.83, pp.853-863, 2006
- [10] T. H. Hyde, M. Saber, and W. Sun, “Testing and modelling of creep crack growth in compact tension specimens from a P91 weld at 650°C,” *Engineering Fracture Mechanics*, vol.77, pp. 2946-2957, 2010,
- [11] I. J. Perrin, and D. R. Hayhurst, “Creep constitutive equations for a 0.5Cr0.5Mo0.25V ferritic steel in the temperature range 600-675°C,” *Journal of Strain Analysis*, vol.31, pp.299-313, 1996.
- [12] A. S. Krausz, and K. Krausz, *Unified Constitutive Laws of Plastic Deformation*. London: Academic Press, 1996.
- [13] M. Wright, *Assessment of Advanced Creep Damage Constitutive Equations*, BEng final year project thesis, 2011, School of Science and Engineering, Teesside University, 2011.

# Short-Term Load Forecasting System Using Data Mining

Liu Jin Yu Jilai

Department of Electrical Engineering  
Harbin Institute of Technology  
Harbin 150001, China  
Email: liujin@hit.edu.cn

**Abstract**—In this paper, by means of data mining techniques, a platform of data warehouse is designed after preprocessing the huge amounts original data of power system, and a system for short term load forecasting (STLF) is developed, in which there is the synthetic technology of both fuzzy clustering and robust regression model in the platform. The useful data excavated from large amounts of data can offer the effective and accurate load forecasting information for reliable and economic operation of power systems. The validity of the designed system for STLF is shown by the simulation results of an actual power system in China.

**Keywords**- data mining, data warehouse, fuzzy clustering, robust regression, load forecasting

## I. INTRODUCTION

Load forecasting is an important element for economically efficient operation and for effective control of power systems [1]. Period of forecasting in demand can change from hour or day for short-term forecasts to week or year for medium and long-term forecasts respectively. Short-term load forecast (STLF) is a key issue for operation of both regulated power systems and electricity markets. Many operating decisions are based on STLF, such as dispatch scheduling of generating capacity, reliability analysis, security assessment and maintenance plan for the generators [2]. With the rise of deregulation and free competition of the electric power industry in many countries around the world, load forecasting becomes more important than ever before. Load forecasts are vital for the energy transactions in competitive electricity markets [3]. The importance of accurate load forecasts will increase in the future because of the dramatic changes occurring in the structure of the utility industry due to deregulation and competition. The load has complex and non-linear relationships with several factors such as climatic conditions, past usage patterns, the day of the week, and the time of the day. Modeling such relationships with conventional techniques such as time series, regression analysis, ARIMA models, Kalman filtering models and others has been attempted before with varying degrees of success[4]-[6].

Data mining is a process for extracting hidden knowledge from large amounts of data stored either in databases, data warehouses. The process focuses on

finding interesting patterns that can be interpreted as useful information [7]. Data mining techniques can be used to discover untapped patterns of data that enable the creation of new information and make use of technologies such as data warehouse, statistics, pattern analysis, data visual, and artificial intelligence (AI)[8]-[9]. The results of data mining will be applied to the forecast of future development trend.

The power system's databases are susceptible to noise and miss of data due to their huge size. Data pre-processing techniques can improve the quality of the data and the accuracy and efficiency of the mining process. It is thus an important step in the data mining process [10]. It considers only useful data with new meaningful variables, and includes data cleaning, data aggregating, data integration and data normalization.

The main purpose of STLF is to accurately predict the loads for the next day. Here in this paper, we study and design a system for STLF based on data mining technique. The structure chart of the system is shown in Fig.1. The advantage of this structure distinguishes application DBMS. Data warehouse is emphasized particularly on storing and managing topic-oriented data, and obtaining useful information for decision-making from it. The system mainly contains five modules:

- Raw data: the huge amounts of unprocessed original data are analyzed by using query techniques.
- A data warehouse with the snowflake schema is designed for STLF, and useful data are loaded into the warehouse by data pre-processing techniques.
- According to the weekdays and weekends of each solar term, a load pattern is established by applying fuzzy clustering within the data warehouse.
- The forecasting method is given based on robust regression model.
- Forecasting result: this is decision layer on the top, which is the graphical interface of orienting to user's needs for STLF. It can conveniently and rapidly query the forecasting results.

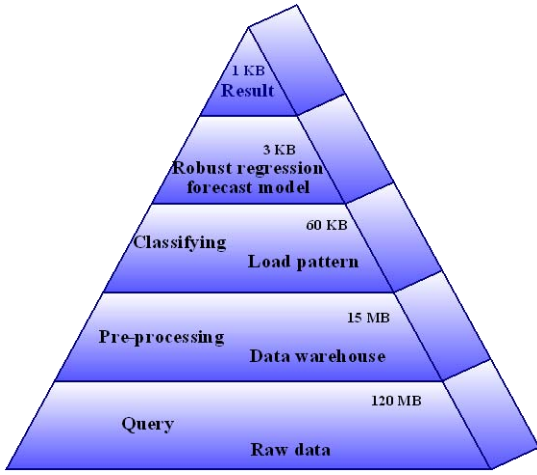


Fig.1 Structure chart of STLF system based on data mining.

## II. DATA PREPROCESSING

### A. Raw data

In the power system, huge amounts of raw data are collected through Supervisory control and data acquisition (SCADA). There are four kinds of data for STLF:

- The data of daily historical load. These data are sent to the host computer server of power system via RTU (Remote Terminal Unit).
- The data of daily history about weather situation gained from observatory via computer network.
- The data for weather forecast such as the highest and lowest temperature and weather type in the forecasted day.
- The data of parameters such as substation, region where a power system locates, etc.

In fact, the above data are the raw data from different data sources, which totals to 120 MB of data each year. Obviously, not all these data are related to the chosen load - some of them have useful information and should be exploited for forecasting, while others may not be correlative with the forecast. In this way, the various isolated information can be concentrated as a whole database, that is, a data warehouse to improve prediction accuracy for STLF.

### C

### B. Data Cleaning and Aggregating

Data cleaning is performed as a pre-processing step when preparing the data for a data warehouse. The raw data of power system are incomplete, noisy, and inconsistent. Data cleaning process works to “clean” the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and correcting inconsistencies.

Considering the requirement of load forecasting, the daily loads should consist of  $n$  ( $n=288, 144, 96$ ) points with regular time intervals. The length of the time interval is  $k$  ( $k=5, 10, 15$ ) minutes. But the raw data of daily historical load consist of 1440 points with 1 minute time intervals. Therefore, the daily load data of the 1

minute time interval are aggregated so as to compute the data of the time interval of  $k$  minute. It can be computed by

$$P_t^* = \frac{1}{s} \sum_{j=t-i}^{t+l} P_j \quad (1)$$

If  $k$  is odd, then  $i=l=(k-1)/2$ , otherwise  $i=k/2-1$ ,  $l=k/2$ .

$P_t^*$  is the load power at time  $t$ ,  $s$  is the nonzero number of the load powers  $P_j$ . Using the above method, not only the daily load data are aggregated, but the noisy data also are smooth.

### C. Data Normalization

Normalization is particularly useful for classification algorithms involving nearest neighbor classification and clustering. In the data warehouse of load forecasting, some attributes need to be transformed into forms appropriate for fuzzy clustering. These attributes are normalized by scaling their values so that they lie within a small specified range, such as 0 to 1. We define some attributes in the weather and date dimension table as a data format:

$$D_k = (D_{k1}, D_{k2}, D_{k3}, D_{k4}, D_{k5}, D_{k6}) \quad (2)$$

In (2),  $k=1, \dots, p$ ,  $p$  is the number of weekdays or weekends during each solar term or every two solar terms;  $D_{k1}, D_{k2}$  are the highest and the lowest temperature respectively on the  $k^{\text{th}}$  day;  $D_{k3}$  is the weather type such as sunny day, rainy day, cloudy day etc.;  $D_{k4}$  is the humidity of the  $k^{\text{th}}$  day;  $D_{k5}$  denotes the type of the  $k^{\text{th}}$  day (either a weekday or a weekend);  $D_{k6}$  denotes which solar term the  $k^{\text{th}}$  day belongs to.

Min-max normalization performs a linear transformation on the original data. Suppose  $D_l^{(\min)}$ ,  $D_l^{(\max)}$  are the minimum and maximum values of  $D_{kl}$  ( $k=1, 2, \dots, p$ ;  $l=1, 2, \dots, 6$ ). Min-max normalization maps a value of  $D_{kl}$  to  $D'_{kl}$ , which lies in the range  $[0, 1]$  by the following relations

$$D'_k = (D'_{k1}, D'_{k2}, D'_{k3}, D'_{k4}, D'_{k5}, D'_{k6})$$

$$D'_{kl} = \frac{D_{kl} - D_l^{(\min)}}{D_l^{(\max)} - D_l^{(\min)}} \quad (3)$$

Min-max normalization preserves the mutual relationships among the original data values.

### D. Data warehouse with snowflake schema

Data integration is performed as a pre-processing step when preparing the data for a data warehouse, which combines data from multiple sources into the data warehouse. These data sources include multiple databases such as the historical load, the historical weather, the forecasted weather, and the parameters of the power system database.

Data warehouse emphasizes particularly on storing and managing theme-oriented data and provides a steady platform for the processing of historical data [11]. The important missions of data warehouse include the creation of a structure for extracting the useful data. The snowflake schema has been recognized as an effective

structure for organizing data warehouse components, which is also a multi-dimensional model composed of a theme-oriented central fact table and a set of surrounding dimension tables [12]. Each dimension table corresponds to one of the components of the fact table.

This paper applies the snowflake schema within the design of the data warehouse for STLF. The structure of the data warehouse is divided into two sorts of table: one is fact table (Load Fact Table), which is used for storing the measurement values (load power, consumed cost) of the facts and the key attribute of each dimension table; the other is dimension table (Time, Date, Location, Region), which is used for storing the describing information, including the layers and member types of the dimension. This structure is easy to maintain and can economize the storage space. A snowflake schema for data warehouse of STLF is shown in Fig. 2.

The fact table is composed of load power and the primary keys of related dimensions tables, such as time dimension, location dimension, and weather dimension. Each dimension table has its own property and is connected with fact table via the key attributes. The data stored into the data warehouse after data pre-processing take up a storage space of 15 MB.

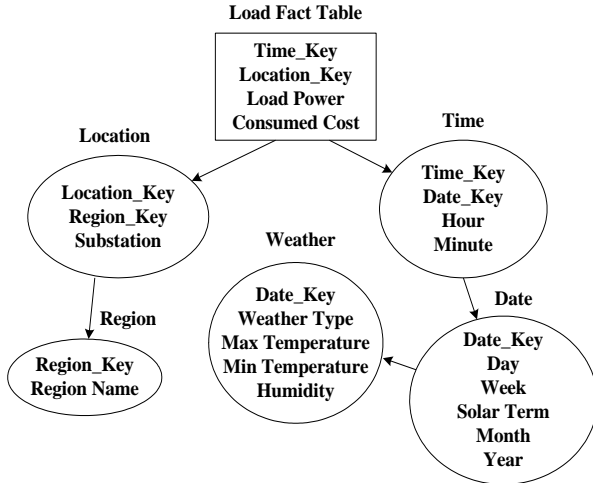


Fig. 2 A snowflake schema

### III. ESTABLISHING LOAD PATTERN

The variety of load is affected by solar-terms. There are twenty-four solar-terms in a year such as “Spring Equinox”, “Great Heat”, “Start of Autumn” and “Midwinter”, etc. In weekdays of a solar-term or the weekends of two solar-terms, there are obvious recurrent characteristics and commonness within the loads curves in power system, i.e. these curves belong to the same load pattern. Therefore, the load pattern should be grasped in order to improve the accuracy of STLF system.

The characteristics of the load curves in two or more adjacent weekdays or weekends are closely related. The load curves of weekdays or weekends are classified into some load patterns in terms of 24 solar-terms of each year. Here in this paper, by using fuzzy clustering method, a load pattern of weekday are established from each solar-term (semi-monthly), a load pattern of weekend are

established from every two solar-terms (monthly). In this way, the load curves belonging to the same load pattern are identified, and the data of these curves are used as the input vector of robust forecast model for the next section.

The fuzzy clustering method identifies natural groupings of data through iteration based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by the membership grade of data points [13]. The method is formulated as an optimization of the objective function

$$z(\tilde{U}) = \sum_{c=1}^2 \sum_{k=1}^p (\mu_{ck})^2 \|D'_k - v^c\| \quad (4)$$

The above function is minimized, subject to the following conditions:

$$v^c = \frac{\sum_{k=1}^p \mu_{ck}^2 D'_k}{\sum_{k=1}^p \mu_{ck}^2}, \quad c = 1, 2 \quad (5)$$

In the above conditions,  $v^c$  is the cluster center of class  $c$  and  $\tilde{U} = [\mu_{ck}]$  is the membership function matrix. The value of  $\mu_{ck}$  can be calculated by

$$\mu_{ck} = \frac{1}{\sum_{c=1}^2 \frac{\|D'_k - v^c\|^2}{\|D'_k - v^c\|^2}}, \quad c = 1, 2, \quad k = 1, \dots, p \quad (6)$$

The load patterns are composed of load curves gained by using the above fuzzy clustering method. Min-max normalization maps a value of  $D_k$  to  $D'_k$  by applying (2) and (3). Then, the group of weather data including the forecasted day is determined by applying (4) to (6). The corresponding load curves are selected from the group as a load pattern:

$$M = (X_1(t), X_2(t), \dots, X_m(t)) \quad (7)$$

$X_1(t), X_2(t), \dots, X_{m-1}(t)$  in  $M$  can be used as the input vector for robust regression model in the following section. They are load curves of the same load pattern, excluding the loads of the forecasted day.

### IV. ROBUST REGRESSION FORECAST MODEL

In load forecasting, outliers are more likely to appear when the load rises rapidly to the peak due to the influence of weather. Some data are difficult to fit with the selected forecasting model, which may directly affect the forecast accuracy. Robust regression model can reduce the influence of outliers through the special weight treatment method. It remedies the limitation of the conventional regression method, which suffers much influence from outliers [13]-[14].

The robust regression models are widely applied to many fields such as economy, biochemistry and computer science [17] -[20]. The robust regression model can be expressed as follows:

$$Y = X\beta + \varepsilon \quad (8)$$

where for  $i=1, \dots, n, j=1, \dots, m, X = (x_{ij})$  is an  $n$  by  $m$



matrix of input vectors of the model,  $X_j(t_i)$  may be obtained  $n$  point observed historical load sample data of the  $j^{\text{th}}$  day, they belong to the same load pattern. The vector  $\beta=(\beta_j)$  can be determined by the method of robust regression.  $\varepsilon=(\varepsilon_i)$  denotes an  $n$  by 1 vector of errors.

The key of determining the prediction model is choosing statistical method to find an ideal estimation vector  $\hat{\beta}=(\hat{\beta}_j)$  from the vector  $\beta$ . Robust regression is to minimize the model error in unusual cases, namely,

$$D(\hat{\beta}) = \min D(\beta) = \min \sum_{i=1}^n \rho(r_i) \quad (9)$$

where  $\rho(\chi)$  is the optimal estimation function, the remainder error  $r_i = y_i - \sum_{j=1}^m x_{ij}\beta_j$ . When the remainder errors obey to normal distribution, the optimal estimation function becomes  $\rho(\chi) = \chi^2$ . In this case, it corresponds to the ordinary least square estimation.

Through theory proving [21], the solution of (9) can be computed by

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \quad (10)$$

Therefore,  $\beta$  in (8) is calculated by using (10), which is the vector of weighted least square estimation.

The key of calculating (10) is to choose a weighting matrix. Here the weighting matrix  $W$  is chosen as

$$W_j(r_j) = \frac{1}{1 + r_j^2} \quad (11)$$

For  $j=1, 2, \dots, m$ ,

$$r_j = \frac{\text{resid}}{s \times \sqrt{1 - h_{jj}}} \quad (12)$$

where *resid* is the vector of residuals from the previous iteration,  $h_{jj}$  expresses a main diagonal element of  $H$ ,  $H = (h_{jj}) = X(X^T X)^{-1} X^T$  as cap matrix which stands for the vector of leverage values from a least squares estimation [22].

For the purpose of robustness, the standard deviation estimation of the error  $s$  [23] is introduced as the following form:

$$s = \text{median} \{ |r_i| \} / 0.6745 \quad (13)$$

The forecasting load model may be given by the following equation

$$\hat{Y}(t) = C + \sum_{j=1}^m \hat{\beta}_j X_j(t) \quad (14)$$

## V. FORECASTING ALGORITHM

According to the load characteristics, an algorithm is proposed for STLFL, which combines the robust regression model with the fuzzy clustering. The flowchart of this algorithm is shown in Fig. 3.

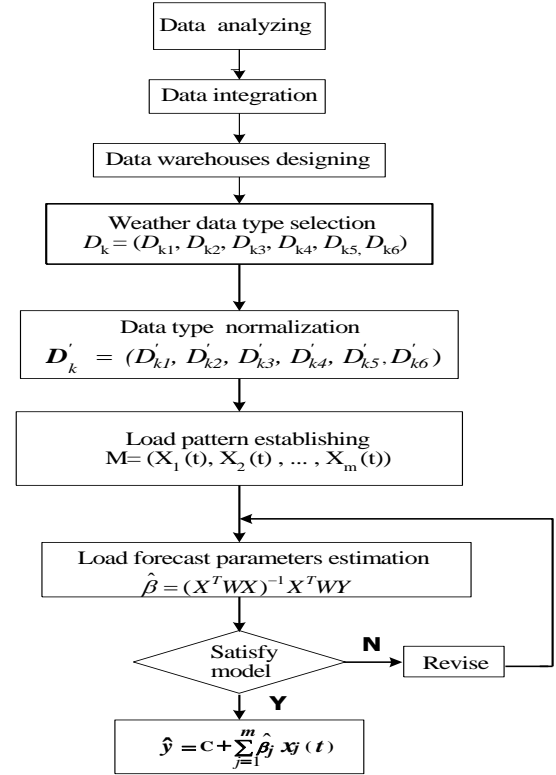


Fig. 3 Flow chart of the algorithm for STLFL

Generally, the loads consist of  $n$  load points with regular time intervals. The length of time interval between every two points is 5 minutes. The algorithm of STLFL is carried out as following steps:

*Step 1.* The huge amounts of raw data are analyzed and pre-processed. Then the data after pre-processing are loaded into data warehouse with the snowflake schema.

*Step 2.* In terms of the type of a day(weekday or weekend), the weather data type  $D_k$  of  $p$  days in a solar term or two solar terms are selected like (2) from the data warehouse, and then Min-max normalization maps  $D_k$  to  $D'_k$  by using (3).

*Step 3.* The weather data are clustered into two groups using (4) to (6), one of which includes the forecasted day. The corresponding load curves are selected from the group as a load pattern  $M$  like (7).

*Step 4.* The output vector  $Y$  and the input matrix  $X$  of the robust regression model are obtained from (8).  $Y=X_m(t)$  is the load curve from the nearest day to forecasting day, and  $X=(X_1(t), X_2(t), \dots, X_{m-1}(t))$  is the other load curves in the same pattern. The estimation parameters  $\hat{\beta}_j$  and the weighting matrix  $W$  of the model are respectively computed by using (10) to (12). Then the model is evaluated. If the model is found to be unsatisfying, it needs to be revised, i.e., the weighting matrix  $W$  can be updated.

*Step 5.* Finally, the output  $\hat{Y}(t)$  of robust regression model is obtained by using (14) as the forecasting load curve.

## VI. SIMULATION RESULTS

The STLF system developed by means of the data mining is tested through an actual power system of China. The daily loads are predicted via the system simulated. The results are shown in Fig. 4 to Fig. 5.

Fig. 4 shows that the weather data are divided into tow groups, in which the group of red dots belongs to the same load pattern. This pattern is shown as follows:

$$X=(X_1(t), X_2(t), \dots, X_6(t))$$

In the pattern,  $X_1(t)$ ,  $X_2(t)$ ,  $X_3(t)$ ,  $X_4(t)$ ,  $X_5(t)$  express the load curves of June 23, June 26, June 27, June 29, June 30 respectively.  $Y=X_6(t)$  is the load curve of July 2.  $X=(X_1(t), X_2(t), \dots, X_5(t))$  and  $Y$  are the input and output vector of robust regression model, respectively.

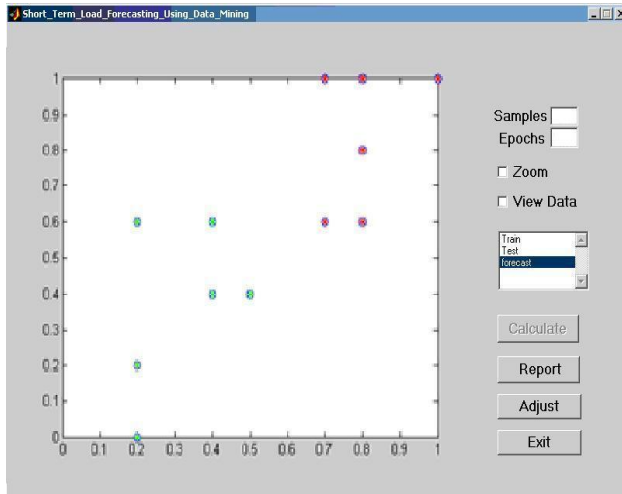


Fig. 4. Fuzzy clustering to a weather data type

The load curve is forecasted on July 2. Two load curves are drawn in Fig. 5, the first is the actual load curve, and the second is the forecasted load curve.

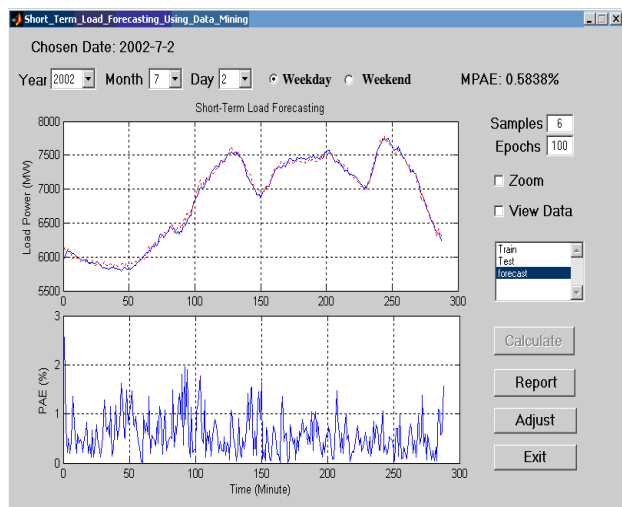


Fig. 5 Actual and forecasted load curves on July 2

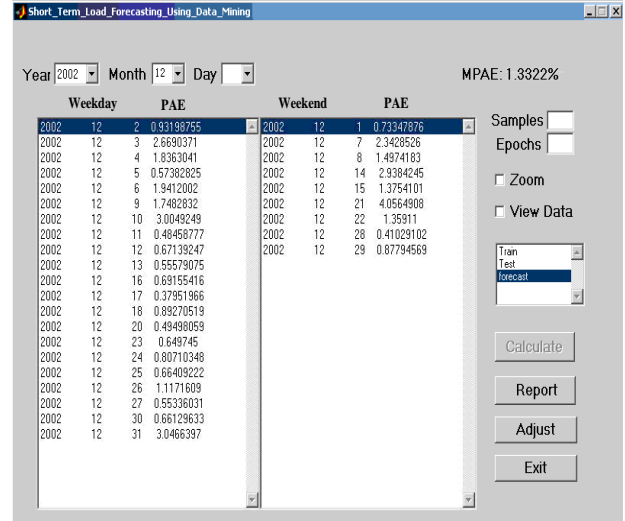


Fig. 6 PAEs of forecasted loads during in a half month

In order to show the effectiveness of the robust regression (RR) method based on data mining, the forecasted accuracy by using the RR is compared with the one adopting the conventional regression (CR) method. Table 1 shows in percentage the absolute errors (PAEs) of forecasted loads using the two methods during a half month in a year. In Table 1, 'Mean of PAE' stands for the mean of daily PAEs of the predicted loads; 'Max PAE' stands for the maximum of daily PAEs during the half month. From 'Table1', it is found out that the precision of load forecasted using RR based on data mining is obviously better than that of the traditional regression method.

TABLE I. PAE OF PREDICTED LOADS FOR CR AND RR

Date	Mean PAE (CR)	Max PAE (CR)	Mean PAE (RR)	Max PAE (RR)
<b>Workday</b>	(%)	(%)	(%)	(%)
6-21	1.2822	4.0713	0.6890	2.1364
6-24	1.6011	5.6684	1.5350	5.2400
6-25	1.4820	3.5794	1.0486	3.1782
6-26	1.3861	2.6072	0.8748	2.2799
6-27	0.9418	2.0959	0.9038	2.0946
6-28	0.8012	2.9947	0.7102	1.7035
7-01	1.2401	2.9203	0.8814	2.5139
7-02	0.6112	1.3234	0.4390	0.8716
7-03	0.9388	1.8870	0.9123	1.8547
7-04	0.5575	1.3219	0.4727	1.1266
7-05	0.7692	1.8771	0.7263	1.8055
<b>Holiday</b>				
6-22	0.7734	2.2071	0.7207	2.1233
6-29	1.0884	2.3151	0.8913	2.3097
6-30	1.0134	2.0461	0.8355	2.0193
7-06	0.7065	2.3422	0.6796	2.3051
7-07	1.4404	3.2036	1.0512	2.4447
<b>Average</b>	1.0396	2.6538	0.8357	2.2504

## VII. CONCLUSION

In this paper, a new STLF system is developed using data mining technique. Firstly, the raw data are collected from different data sources in the power system into data warehouse after data pre-processing, then the load patterns are established from the data warehouse by applying fuzzy clustering in terms of the condition of weather, weekdays or weekends. There is a connection between the load curves of the past days and the load curve of the forecasting day by using robust regression. The simulation results show that the proposed forecasting models exhibit very good forecast qualities for both weekdays and weekends. Using the integration technique such as data warehouse, fuzzy clustering, robust regression, the useful information mined from the data warehouse can provide effective and exact forecasting loads for managers and dispatchers of power system.

## REFERENCES

- [1] R. Mamlook, O. Badran, E. Abdulhadi, "A fuzzy inference model for short-term load forecasting" in *Energy Policy*, 2009, vol. 37(4), pp.1239-1248.
- [2] N. Amjady, F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm", *Energy*, 2009, vol. 34(1): 46-57.
- [3] F. L. Chen, "Short-term load forecasting based on an adaptive hybrid method", *IEEE Trans Power Syst*, 2006, vol. 21 (1), pp. 392-401.
- [4] M. Telbany, E. Karimi F. ELECTRIC, "Short-term forecasting of Jordanian electricity demand using particle swarm optimization", *Power Systems Research*, 2008, vol. 78(1), pp. 425-433.
- [5] H. M. Hamadi, S. A. Soliman, "Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model", *Power Systems Research*, 2004, vol. 68(2), pp. 47-59.
- [6] S. E. Papadakis, J. B. Theocharis, A. G. Bakirtzis, "A load curve based fuzzy modeling technique for short-term load forecasting", *Fuzzy Sets and Systems*, 2003, vol. 135, pp. 279-303.
- [7] M. Cannataro, D. Talia, P. Trunfio, "Distributed data mining on the grid", *Future Generation Computer Systems*, 2002, vol. 18, pp. 1101-1112.
- [8] R. Nemati, M. Steigerb, L. S. Iyerc, R. T. Herschel, "Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing", *Decision Support Systems*, 2002, vol. 33, pp. 143- 161.
- [9] Z. H. Zhou. Three perspectives of data mining, "Artificial Intelligence", 2003, vol. 143, pp. 139-146.
- [10] J. Han, M. Kamber. Data mining: Concepts and Techniques. Beijing: China Machine Press, 2001.
- [11] D. Theodoratos, T. Sellis, "Designing data warehouses," *Data & Knowledge Engineering*, 1999, vol. 31, pp. 279-301.
- [12] M. Levene, G. Loizou, "Why is the snowflake schema a good data warehouse design," *Information Systems*, 2003, vol. 28, pp. 225-240.
- [13] F. S. Cheng, W. D. Wei. Fuzzy Mathematics and Fuzzy Optimizing, Beijing: Science Press, 1997, pp. 309-318.
- [14] F.K. Wang, Y.F. Cheng, "Robust regression for estimating the Burr XII parameters with outliers," *Journal of Applied Statistics*, 2010, vol. 37(5), pp. 807-819.
- [15] X .Bao, L.K. Dai, "Iterative robust least square support vector machine for spectral analysis," *Asian Journal of Chemistry*, 2010, vol. 22 (6), pp. 4511-4523.
- [16] A. Zaman, P. J. Rousseeuw, M. Orhan, "Econometric applications of high-breakdown robust regression techniques," *Economics Letters*, 2001, vol. 71, pp. 1-8.
- [17] E. B. David, L. Kidong, "Robust regression-based analysis of drug-nucleic acid binding", *Analytical Biochemistry*, 2003, vol. 319, pp. 258-262.
- [18] A. C. Atkinson, C. C. Tsung, "On robust linear regression with incomplete data," *Computational Statistics & Data Analysis*, 2000, vol. 33, pp.361-380.
- [19] W.T. Cui, X.F. Yan., "Adaptive weighted least square support vector machine regression integrated with outlier detection and its application in QSAR," *Chemometrics And Intelligent Laboratory Systems*, 2009, vol. 98(2), pp. 130-135.
- [20] A. Giloni, P. M. Least, "Trimmed squares regression, least median squares regression, and mathematical programming," *Mathematical and Computer Modeling*, 2002, vol. 35, pp. 1043-1060.
- [21] J. Liu, Y. Pan , T. X. Long, and . J. L. Yu, " Research on peak load forecasting based on the strategy for preserving steep rise information of front position," *Proceedings of the CSEE*, 2004, vol.4(5), pp. 12-17.
- [22] H. X. Qun. Regression analysis and application. Beijing, Press of the People University of China, Sep. 2002, pp. 59-63.
- [23] H. Edelgard, M. Luc, S. V. Johanna, " Robust regression and outlier detection in the evaluation of robustness tests with different experimental designs," *Analytica Chimica Acta*, 2002, vol. 463, pp. 53-73.

# Applying the Design of Experiment (DoF) to Optimise the NN Architecture in the Car Body Design System

Sugiono\*, Mian Hong Wu, Ilias Oraifige

Dept. of Arts, Design & Technology

University of Derby, Derby, United Kingdom

Sugiono\_ub@yahoo.com, m.h.wu@derby.ac.uk, i.oraifige@derby.ac.uk

**Abstract** — Neural Network (NN) architecture is very important part to establish the best performance of NN. As a consequence, a lot of investments have been done in this research area. This paper is going to show how the design of experiment (Taguchi method) selects the neural network parameters in car body design system. NN architectures included dealing ways with number of neurons, number of hidden layers, transfer functions, learning algorithms and factors interaction. The paper employed Genetic algorithm (GA) which is built in function of software to adjust learning rate, momentum, additive, multiplicative and smoothing. Finally, the NN modules will be used in car body design system to provide the information of external aerodynamic noise, aerodynamic vibration and fuel consumption factors for the user or car body designer.

*Key words; Neural Network, Genetic Algorithm, Taguchi Method, car body design*

## I. INTRODUCTION

Back propagation neural network (BPNN) are highly applicable for many problems, e.g. industry, health, military, financial, etc. A Neural Network can be described as a black box that knows how to process inputs system to create an useful output as a goal or target with the calculation is very complex and difficult to understand by using mathematical model. Neural network copied the working system of biological nervous system as example the brain for processing the information. Akram A. on his journal defined a neural network as The neural network model is a data structure that can be adjusted to produce a mapping from a given set of input data to features of or relationships among the data. The model is adjusted, or trained, using a collection of data from a given source as input, typically referred to as the training set [1]. The problem of identifying the topology of neural network architecture is an important subject to produce a high quality of BPNN performance as is indicated by the lowest means square error (MSE), stabilize the performance and the shortest time-consuming during the NN training. A commonly approach in establishing the NN parameters is through the trial and error method. This method is an inferior quality neural network with increasing resources of cost, time and energy. The optimising the number of hidden layers and the number of neurons for a FNN (feedforward neural network) to solve a practical problem remains one unsolved problem in this research area. The current studied only considered the input dimension and the number of

training pairs in the data set. A good idea for this research would be also considered the other factors which can affects the performance of hidden layer.

The collection of aerodynamic data for noise, vibration and fuel consumption factors of car body design can be very expensive in real car test with requirements of object (cars model), test facility, personal skill and energy demand. Generating numerical data by using computational fluid dynamic (CFD), finite element analysis (FEA) and computational aeroacoustic (CAA) are to be the best option to collect the data with good quality validation and testing performance. Environment effects of increasing the amount of vehicles on the streets and improving human mobility demand are the main reasons it needs investigation of fuel consuming factors, aerodynamic noise and aerodynamic effect in car body vibration. The aerodynamic noise will become more important than other noise (machine noise and tyre noise) when the speed of car is over 80 km/h [2]. Roberts Bosch on his book of automotive handbook 5<sup>th</sup> reported that the reducing of aerodynamics drag is a vital object for improving fuel economy and for achieving the best speed with recognized power machine [3]. It emphasises that the drag resistance is an important parameter to build an economic car body design. The low frequency vibration (less than 0.5 Hz) as mentioned in ISO 2631 can cause pallor, sweating, nausea and vomiting [4].

Three types of passenger cars (saloon, estate and hatchback) are used to verify the NN modules. The inputs system are important geometry parameters to describe the car body aerodynamic behaviors and the outputs system describe an external aerodynamic noise, fuel consumption factors (drag coefficient and Lift/down force), and car body vibration. Delivering NN modules successfully leaved conventional approach for CFD, FEA and CAA simulation with reducing time-consuming, energy and research cost.

## II. METHODOLOGY

The study area of the paper is developing a strong performance of NN with selecting the best parameter s level. The best NN architecture is used to train the database with the information of fuel consumption factor, noise and vibration in car body design. The project concerned in three types of hatchback, saloon and estate car passenger.

### A. Building Car Body Databases

Shapes car body design are developed by identifying the important parameters of car body dimension for saloon, estate and hatchback types. The combinations of important parameters are used to develop the car body design in CAD software and then it is tested by using CFD, CAA and FEA to get the information of noise, vibration and fuel consumption. All the information are collected and putted them in database input-output car body design. The database will be trained by BPNN robust design to produce NN modules to the user. The main activity of generating shapes car body database is transparently demonstrated in fig. 1.

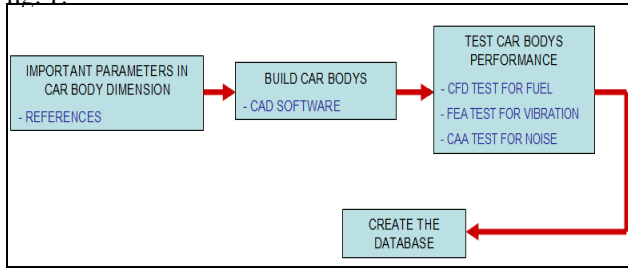


Figure 1. Steps to build the database in shapes car body design

The important parameters of each car type are defined by the sensitivity of model in pressure configuration for fuel consumption and the sensitivity of model in vortices configuration for aerodynamic noise performance. Figure 2 demonstrates one example of important parameters to define the car body shape for front end angle ( ) which has been revealed by many researchers e.g. book of aerodynamic of road vehicle by Wolf Heinrich Hucho [5].

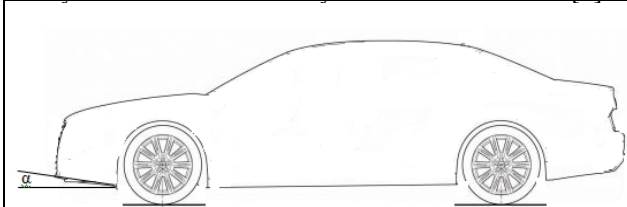


Figure 2. Example of important parameters for the front end angle car ( ).

All the important parameters for saloon car are displayed in table 1 with eleven shape factors and one speed factor. Both of estate and hatchback car type are demonstrated in five shape factors and one speed factor. Levels or components of variable are depended on sensitivity geometry for CFD and CAA test. References are also used to define some parameter s levels.

TABLE 1. GEOMETRIES PARAMETERS FOR IDENTIFYING SALOON CAR

var.y	Var.α	Var.γ	Var.β	Var.θ	Var.δ	Var.σ	Var. Lr/lar	Var. Lh	Var. Rf/W	Var. Lb/hr	Var. v
152.00	0.00	5.00	0.00	45.00	15.00	0.00	0.02	1200	0.25	0.25	25
160.00	2.50	7.50	2.50	55.00	20.00	5.00	0.04	1300	0.35	0.50	50
170.00	5.00	10.00	5.00	65.00	25.00	10.00	0.06	1400	0.40	0.75	75
	10.00	12.50	7.50		30.00	15.00	0.08	1500	0.45	1.00	100
	15.00	15.00	10.00		35.00	20.00	0.10	1600	0.50	1.25	125

Levels of saloon car in 12 variables (see table 1) and levels of estate or hatchback car in 6 variables are combined by using random generation method to capture

good distribution combination. The result of levels and variables combinations are classified as input data for experiment treatment. All treatments are tested by using CFD, CAA and FEA to deliver the output targets. Saloon car has 100 total treatments and both of estate and hatchback car type have 75 total treatments. Table II shows the example of database inputs and outputs variables for saloon car body design.

TABLE II. INPUTS OUTPUTS DATABASE FOR SALOON CAR BODY DESIGN

ID	SALOON CAR'S PARAMETERS INPUTS										
	var. y	Var. α	Var. γ	Var. β	Var. θ	Var. δ	Var. σ	Var. Lr/lar	Var. Lh	Var. Rf/W	Var. Lb/hr
1	152	0.00	15.00	0.00	65	30	0	0.02	1200	0.25	0.25
2	152	0.00	12.50	2.50	45	35	5	0.04	1300	0.35	0.50
3	152	0.00	5.00	5.00	55	15	10	0.06	1400	0.40	0.75
4	152	0.00	7.50	7.50	65	20	15	0.08	1500	0.45	1.00
5	152	0.00	10.00	10.00	55	25	20	0.10	1600	0.50	1.25
6	152	2.50	15.00	2.50	55	20	20	0.02	1300	0.40	1.00
7	152	2.50	12.50	5.00	65	25	0	0.04	1400	0.45	1.25
8	152	2.50	5.00	7.50	55	30	5	0.06	1500	0.50	0.25
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-

Noise (Db)	Fuel Consumption Factors		Aerodynamic Vibration (Hz)				
	Cd	Fl (N)	Z	Y	YZ	X	XY
78.60	0.70940	159.55	80.71	182.06	271.03	274.86	296.41
84.60	0.51719	222.65	73.37	176.58	179.98	236.62	275.39
80.80	0.46395	422.56	71.29	147.24	166.57	185.35	239.96
99.40	0.45190	401.10	67.18	111.34	156.80	164.60	252.96
95.30	0.46595	41.20	43.24	99.77	142.44	154.21	233.07
93.40	0.56179	431.32	75.01	103.89	152.34	188.19	244.24
76.50	0.46642	151.96	70.93	83.54	134.45	177.80	218.20
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-

### B. Calculating for Fuel Consuming Factors, Aerodynamic Noise and Aerodynamic Vibration

First step to deliver the database is understanding how to calculate the information of fuel consumption factors (drag coefficient factor and lift/down force factor), external aerodynamic noise through the car body design and effect of external aerodynamic in car body vibration. Some assumption, approaching and validation are addressed to describe the simulation system as look like the real environment condition.

#### 1) Fuel consuming factors

Roberts Bosch on his book of automotive handbook 5<sup>th</sup> presented the fuel consumption formula is dominated by drag coefficient (Cd) and lift or down force (FL) as aerodynamic effect through the car body design (see formula 1). Drag coefficient is built by dividing the aerodynamic resistance force to parameters of density, fluid speed and frontal area of car body design. Lift force/ down force is developed as effect of deference between fluid pressure above car body and under car body design. The negative pressure coefficient under car body comparing to upper car body will generate down force.

$$Be = \frac{\int_{be} \cdot \frac{1}{\eta_s} \left[ \left( m \cdot f \cdot g \cdot \cos \alpha + \frac{\rho}{2} \cdot c_d \cdot A \cdot v^2 \right) + m(a + g \cdot \sin \alpha) + Br \right] \cdot y \cdot dt}{\int v \cdot dt} \quad .(1)$$

Where:

Be = Consumption per unit of distance (gram/m)  
= Transmission efficiency of drive train (%)

$m$  = Vehicle mass (kg)    Lift force (kg)  
 $f$  = coefficient of rolling resistance  
 $g$  = gravitational acceleration ( $\text{m/s}^2$ )  
 $\theta$  = Angle of ascent ( $^\circ$ )  
 $\rho$  = air density ( $\text{kg/m}^3$ )  
 $c_d$  = Drag Coefficient  
 $A$  = frontal area ( $\text{m}^2$ )  
 $v$  = vehicle speed ( $\text{m/s}$ )

Computational fluid dynamic (CFD) is employed in this session to find out the value of  $C_d$  and FL with varieties of car speed e.g. 25 mph, 50 mph, 75 mph, 100 mph and 125 mph. There are some steps to determine the drag coefficient and lift force in CFD test. First step is creating the car body model in CAD software as an object experiment based on existing car body dimension, as example is VW golf car design in fig. 3. The next step is setting the CFD simulation e.g. defining the  $C_d$  performance in equation form. Finally, monitor the simulation mechanism to reach the convergence condition for delivering the final drag coefficient and lift/ down force performance. The total of time-consuming and total of iteration of CFD simulation is depending on the complexity of the cad design and the PC specifications.

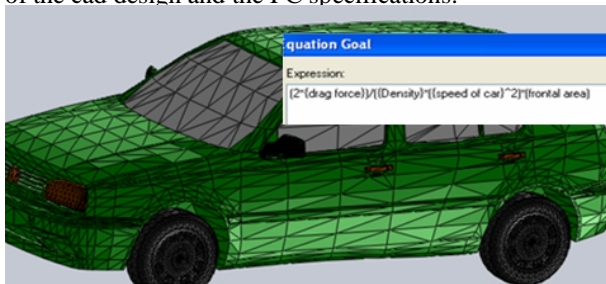


Figure 3. CAD model and drag coefficient formula in CFD.

## 2) External aerodynamic noise

In this paper, acoustics far field analogy method is used to define the sound noise around vehicle body designs as effect of external aerodynamic flow. Ffowcs William Hawkins (FW-H) equation is one of far field acoustic modeling that working in integral method based on Lighthill acoustic analogy offer variable alternatives to the direct method. Moreover, the Ffowcs Williams and Hawkins (FW-H) equation is fundamentally uniform waves equation that is derived by manipulating the continuity equation and the Navier-Stokes equations. Fluent offer FW-H formulation is adequate to predict sound generated by equivalent acoustic sources such as monopole, dipole and quadrupoles. Time accurate solutions of the flow-field variables, such as vortices, pressure, velocity component, and density on source (emission) surface, are important parameters to evaluate the surface integrals. Time-accurate solution is obtained from unsteady Detached Eddy simulation (DES) equation.

Post processing of sound signal from receiver to graph result of frequency sound pressure level (SPL) is conducted by Fast Fourier Transform (FFT) method. There are some steps to solve this problem: first step is setting the noise simulation in Computational Aeroacoustic ANSYS

fluent software, e.g. receiver position in (2, 1.8, 2) coordinate system and car speed in  $\text{m/s}$ . Secondly step is selecting FW-H acoustics model in time step size 0.00001 and number of times 1000. Receiver position here is defined as the highest noise that has produced by inference waves of air flow through the car body design from any possible receiver position (see fig. 4a). Finally, the last step of noise calculating is reading the results in SPL frequency graph for estimating the external aerodynamic noise in any shapes car body design (see fig.4b).

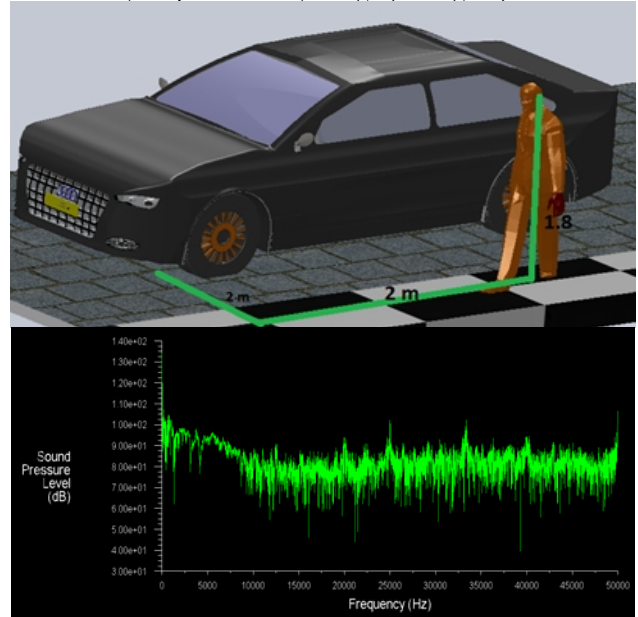


Figure 4. Computational aeroacoustic simulation: a. setting the receiver position and b. SPL frequency graph result.

## C) Aerodynamic Effect in Car Body Vibration

The vibration of car body design as effect of external aerodynamic flow through the car body is investigated by finite element method (FEM) in five modes of direction in x, y, z coordinate position. First step to calculate the vibration is building the car models based on parameters in table I by CAD software. Secondly step is defining the restrain position to hold the car model. Third step is generating the mesh geometry in the best quality based on computer's memories available. Finally, importing the external force from CFD results to the FEA setting and running the numerical simulation test to generate 5 modes of frequencies.

## III. OPTIMIZATION OF NN PARAMETERS

According to the references, the NN topology is a black box condition that it needs more parameters investigation to produce a good NN performance. Design of experiment (DoF) Taguchi method is employed to select the best NN parameters based on the car body databases in limitation data test of CFD, FEA and CAA simulation. The optimal NN's architecture is addressed to train and to test the database for delivering NN modules.



#### A. Defining Orthogonal Parameters for BPNN Problems

There are seven factors and 3 interactions are investigated in Taguchi test e.g. number of hidden layers, transfer function, number of neurons, learning algorithm, etc (see table III). An A factor has the two levels for one hidden layer and two hidden layer. A dummy variable is addressed for replacing A factor at level 3 to be A factor at level 2 with hypothesis that two hidden layers will produce better performance than one hidden layer in BPNN training. The B and C factors-levels are built based on the Kolmogorov s and Lipmann s approaches as shown respectively below which can be used for setting the lower and the upper of the number of neurons in the hidden layers [6]. Symbol N is for total number of inputs and symbol OP is the total variables of output side.

Base on this problem with 12 inputs for saloon car type and 6 inputs for both of estate and hatchback car, the average of inputs total is  $24/3 = 8$  and the variable outputs total is 8. These values are used to configure the neurons in the hidden layer as:

- Number of neurons in the first hidden layer is:  
Lower neurons:  $2N + 1 = 2(8) + 1 = 17$   
Upper neurons:  $OP \times (N + 1) = 8 \times (8+1) = 72$
- Number of neurons in the second hidden layer is:  
Lower neurons:  $2N + 1 + (2N + 1)/3 = 17 + 17/3 = 23$   
Upper neurons:  $OP \times (N + 1) + (OP \times (N + 1))/3 = 72 + 72/3 = 96$

To create the effective design of back propagation neural networks, the research has applied the genetic algorithm which is provided by NN software as one part to optimise the NN parameters. The main advantage of the using GA is associated with its ability to discover

automatically a new value of neural network parameters from initial value. The initial values of learning rate is 0.500000 initial momentum is 0.0166, number of populations is 50 chromosomes, using roulette selection, initial weigh of network factor is 0.1074, working in heuristic mutation with mutation probability is at 0.01, crossover heuristic probability is 0.90 and maximum generation number is 100.

A G factor is provided for three levels of learning algorithms for:

- 1) Conjugate gradient learning algorithm with GA is used to optimise the number of neurons in hidden layer and to optimise the input data configuration.
- 2) Quickprop learning algorithm with GA is to optimise the momentum, learning rate and input data configuration,
- 3) Delta bar delta learning algorithm with GA to optimise the learning rate, additive, multiplicative, smoothing and input data configuration.

Time-consuming and huge memory requirement are the main reason to eliminate the Levenberg Marquard learning algorithm factor.

In the Taguchi method, the appropriate of orthogonal array depended on the particular degrees of freedom (*dof*) of the design experiment. Base on table III which has seven factors in 3 levels and 3 interactions of A X B/C, AX D/E AND B X D/E, the total *dof* can be determined as following this calculation:

$$\begin{aligned} \text{dof} &= 1 + 7(3-1) + 3(3-1)(3-1) \\ &= 27 \end{aligned}$$

The *dof* = 27 is used to choose an adequate orthogonal array  $L_{27}(3^{13})$  to accommodate 7 factors and 3 interactions.

TABLE III. SETTING PARAMETERS OF BPNN TOPOLOGY FOR TAGUCHI TEST

No	BPNN's Factors	Level's Parameters		
		Level I	Level II	Level III
1	Number of hidden layers (A)	1	2	2
2	Number of Neurons-hidden layer I (B)	17*	40	72*
3	Number of Neurons-hidden layer II (C)	23*	50	96*
4	Transfer Function for Input - hidden layer I (D)	Sigmoid	Tanh	Linier
5	Transfer Function for hidden layer I -hidden layer II(E)	Sigmoid	Tanh	Linier
6	Transfer Function for hidden layer - output (F)	Linier Sigmoid	Linier Tanh	Linier
7	Learning Algorithm (G)	Conjugate Gradient + GA	Quick Prop + GA	Delta Bar Delta + GA
8	Epoch (H)	1000	5000	10000
9	Interaction A x B/C			
10	Interaction A x D/E			
11	Interaction B/C x D/E			
12	Error (I)			

\*Number of neurons in hidden layer is based in formula at reference [6]

#### B. Running the BPNN Based on Orthogonal Array Setting

After fill all the BPNN parameters in orthogonal array as inputs experiment matrix, the next step is to find the result of MSE performance in three noise factors (saloon, estate and hatchback types). MSE as a response variable is built by testing the neural network parameters configuration used artificial intelligence software/ program. According to orthogonal array in table IV, the simulation is taken in 27 BPNN testing, e.g. simulation in 15<sup>th</sup> raw have NN configuration of two hidden layer (A factor at 2<sup>nd</sup>

level), 40 neurons in first hidden layer (B factor at 3<sup>rd</sup> level), 50 neurons in second hidden layer (C factor at 2<sup>nd</sup> level), transfer function for inputs layer to hidden layer in linier axon (D factor at 3<sup>rd</sup> level), transfer function for first hidden layer to second hidden layer in Linier axon (E factor at 3<sup>rd</sup> level), transfer function for second hidden layer to outputs layer in Tanh axon (F factor in 2<sup>nd</sup> level), Delta Bar Delta learning (G factor at 3<sup>rd</sup> level) and epoch for 10.000 (H factor at 3<sup>rd</sup> level) produced MSE 0.0984 for saloon car, 0.0237 for hatchback car and 0.0286 for estate car.

### C. Taguchi Test Results for BPNN Optimum

By using the statistic or especially using design of experiment software, the orthogonal array in table IV for saloon car body design is executed to find out the best NN parameters, to investigate the effect for each parameter and to know the interaction for each parameter. An ANOVA, ANOM and S/N ratios measurements are employed together to investigate this problem. Analysis of variance

(ANOVA) was performed for each of the performance measures using the S/N ratio as the response in Taguchi test. The goal of conducting variance analysis with parameters experiment is to determine the relative magnitude of the effect of each factor on the objective functions of MSE performance. The best level of each parameter or factor design is indicated by the highest value of S/N ratios performance.

TABLE IV. ORTHOGONAL ARRAY FOR BPNN OPTIMIZATION IN TAGUCHI TEST

No	BPNN PARAMETERS													BPNN RESULTS		
	A	B/C	A X B/C	NIL	D/E	A X D/E	NIL	B/C X D/E	F	G	NIL	H	e	MSE Saloon	MSE Estate	MSE Hatchback
1	A1	B1	1	1	D1	1	1	1	F1	G1	1	H1	1	0.0214	0.0103	0.0045
2	A1	B1	1	1	D2	2	2	2	F2	G2	2	H2	2	0.0281	0.0218	0.0245
13	A2	C2B2	3	1	D1E1	2	3	2	F3	G1	3	H1	2	0.0302	0.0232	0.0210
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	A2	C3B1	2	1	D2E2	1	3	1	F3	G2	3	H2	1	0.0931	0.0447	0.0210
27	A2	C3B1	2	1	D3E3	2	1	2	F1	G3	1	H3	2	0.0085	0.0098	0.0065

The S/N ratios of Taguchi result is displayed in table V with eight factors experiment, three interactions and an error factor. The interaction factors are between number of hidden layers (A) and number of neurons in hidden layer (B/C), between number of hidden layers (A) and transfer function (D/E), and the interaction between number of neurons (B/C) and transfer function (D/E). The best level for all factors are A factor is in first level or neural network with one hidden layer, B factor is in third level with 96 neurons, D factor is in first level with Sigmoid transfer function, F factor is in first level with linier sigmoid function, G factor is in third level with delta bar delta learning (GA is used to optimise learning rate, momentum, etc.) and H factor is in third level with 10.000 epoch.

TABLE V. S/N RATIOS ANALYSIS FOR THE BEST NN PARAMETERS

Level	A	B	C	A X B/C	D	E	A X D/E	B/C X D/E	F	G	H	E
Level 1	30.54	28.49	28.88	30.39	29.47	28.75	29.16	29.18	37.64	27.96	28.63	28.09
Level 2	28.60	29.18	29.37	28.65	29.43	28.79	29.63	31.72	26.57	29.56	29.24	30.85
Level 3		29.90	27.71	28.84	28.62	28.27	28.72	27.53	28.97	29.73	29.65	28.96
Effect/ Difference	1.94	1.41	1.66	1.74	0.85	0.52	0.92	4.18	11.07	1.77	1.02	2.76
Rank	IV	VIII	VII	VI	XI	XII	X	II	I	V	IX	III
Optimum	A1	B3			D1				F1	G3	H3	

Figure 5 demonstrates the interaction results of A, B and D factors in linier graph. The interaction factors between A (number of hidden layers) and B (number of neurons) is in categorizing low relationship. The same condition is the relationship between A factor and D factor (transfer function from input layer to hidden layer) has a low impact in MSE performance. In contras the strong relationship is occurred in correlation between B factor and D factor.

Fowlkes and Creveling suggested looking at the F-ratios results in the ANOVA table to know the influence of the design experiment factors with the following criteria [7]:

1.  $F_{test} < 1$ : Section effect is insignificant (experimental error outweighs the control factor effect).
2.  $F_{test} = 2$ : Section has only a moderate effect compared with experimental error.
3.  $F_{test} > 4$ : Section has a strong (clearly significant) effect.

According to ANOVA analysis, it is clearly that the most influence factor for building a strong BPNN design is transfer function between hidden layer to output layer or in F factor with contribution  $F_{test} = 18.82$  or around 79.11 % from the total of seven factors. There is one factor should be categorized in significant effect (F Factor) and one factor is set to moderate effecting (ERROR factor). The others factors are placed in insignificant effect (A, B/C, D/E, G and H). The error value could be represented the other parameters such as learning rate, momentum, initial weight, training sample size, interaction each factor, empty columns in orthogonal array, etc.

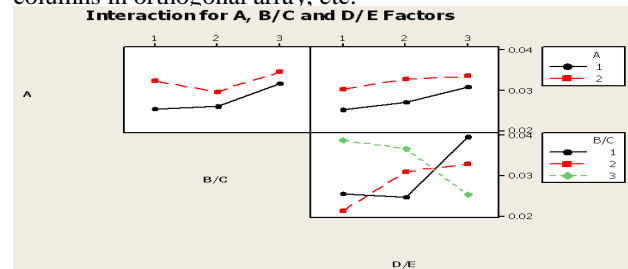


Figure 5. Interaction plots for A x B/C, A x D/E, and B/C x D/E based on S/N ratios

Finally, it can be declared that the best NN parameters are: one hidden layer, 96 neurons in hidden layer, sigmoid transfer function between input to hidden layer, linier sigmoid transfer function between hidden layer to output layer, delta bar delta learning algorithm with optimised by genetic algorithm, epoch 10.000 and strong interaction between number of neurons - transfer function from input to hidden layer.

The next step is to confirm that the recommended best settings of the proposed neural network architectures will establish the performance improvement of NN training and testing. In addition, the MSE under the best settings are comparable to the best of the 27 main experiments from Table IV in terms of both of the networks accuracy and the convergence speed / time-consuming. From these results, we can confidently declare that the resultant neural network using the recommended settings will be significantly robust design for actual applications with average MSE 0.007928872 and 52 minutes time-consuming for saloon car, avg. MSE is 0.008209671 and time-consuming in 43 minutes for hatchback car type and avg. MSE is 0.009366784 and time-consuming in 49 minutes for estate car type.

#### IV. TRAIN, TEST AND VERIFICATION OF NN MODULES

The NN modules provide the form of inputs and form for outputs in excel. The users must initiate the inputs of important parameters to know the information of their design for aerodynamic noise (dB), aerodynamic vibration (Hz) and fuel consuming factors of drag coefficient/ Cd lift force (N) at the same time.

##### A. Train the Databases

The best NN topology is employed to train the databases of saloon, estate and hatchback car type based on the back propagation learning to compare the actual and the desire output which was configured in MSE performance. Figure 6 is an example of training trajectory to reach the noise output closes to noise target in 25 exemplar simulation test of the saloon car. Genetic algorithm (GA) in NN software presented the optimization of inputs configuration, learning rate, momentum, additive, multiplicative and smoothing parameter with selecting heuristic crossover, uniform probability mutation for 0.01 and working in population size 50. The performance of the best fitness of MSE versus generation and average fitness of MSE versus generation are investigated with appearance convergence for producing the final MSE.

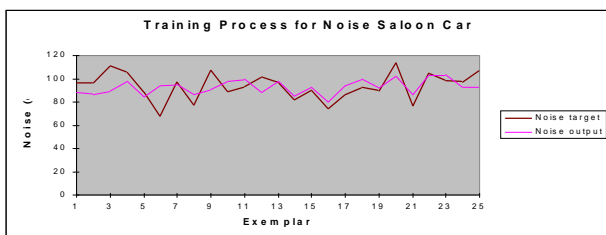


Figure 6. Training trajectory for comparing input - output desire in saloon car database training

##### B. Test & Validation Modules

After the database training has done, the next action is testing the NN module for producing the goal information of noise, lift force, drag coefficient and vibration. Comparing NN module and data test from CFD, CAA and FEA simulation results are used to validate the system as example, saloon Audi car performance for:  $\rho = 8.09^0$ ,  $\mu = 10.05^0$ ,  $\nu = 10.07^0$ ,  $\omega = 65.21^0$ ,  $\phi = 20.64^0$ ,  $\psi = 4.83^0$ ,  $L_h = 1390.1$  mm,  $L_r/ar = 0.007$ ,  $R_w/W = 0.35$ ,  $L_b/L_h = 0.79$ ,  $y$

$= 170$ mm and  $v = 25$  mph has output information NN modules for: noise = 88.6dB, Cd = 0.4042, FI = -49.80, vibration Z = 78.68, Y = 132.02, YZ = 163.85, X = 196.88, and XY = 240.04 and simulation test for: noise = 76.4dB, Cd = 0.40641, FI = -65.80, vibration Z = 86.91, Y = 149.06, YZ = 169.13, X = 182.89, and XY = 241.87.

## II. CONCLUSION

Design of Experiment - Taguchi method is an effective and robust design tool to build the best BPNN architecture from the possible parameters for improving MSE performance in seven factors and three interactions which were explicitly considered robustness as a significant design criterion. The most significant BPNN s factor was investigated with 79.11 % is contributed by transfer function between hidden layer to output layer. The interaction between number of neurons and transfer function in the hidden layer has a powerful relationship to give influence in NN performance. The paper successfully builds the NN modules for Saloon, Estate and Hatchback car including the verification error for all databases to collect the information of fuel consumption factors, aerodynamic noise and aerodynamic vibration without involving conventional approach e.g. CFD, FEA and CAA simulation.

## ACKNOWLEDGMENT

Thank to University of Derby, UK for supporting this paper. The authors are also grateful to the Ministry of National Education of the Republic of Indonesia and University of Brawijaya, Malang Indonesia for their scholarship.

## REFERENCES

- [1]. Akram A. Moustofa, Performance evaluation of artificial neural network for spatial data analysis , Contemporary engineering sciences Vol.4, 2011.
- [2]. Ono K, Himeno Ryutaro, Fukushima Tatsuya, Prediction of wind noise radiated from Passenger cars and its evaluation based on auralization , Elsevier, 1999.
- [3]. Bosch R., Automotive handbook 5<sup>th</sup> , Roberts Bosch GmbH, 2000.
- [4]. Burber, Antony, Hanbook of noise and vibration control , Elsevier advanced tech., UK, 1992.
- [5]. Wolf-Heinrich Hucho, Gino Sovran, Aerodynamic of road vehicle , Annual review of fluid mechanics, Volume 25@1993.
- [6]. G. Ambrogio and L.Filice, Application of neural network technique to predict the formability in incremental forming process , University of Calabria, Italy, 2009.
- [7]. Fowlkes, W. Y., & Creveling, C. M. Engineering methods for robust product design: using Taguchi methods in technology and product development , MA, 1995.
- [8]. Browand Fred, Reducing aerodynamic drag and fuel consumption , Advanced transportation workshop, California, 2005.

# Multiobjective Design of Evolutionary Hybrid Neural Networks

Lavinia Ferariu and Bogdan Burlacu  
Dept. of Automatic Control and Applied Informatics  
“Gheorghe Asachi” Technical University of Iasi  
Iasi, Romania  
lferaru@ac.tuiasi.ro

**Abstract**—The paper presents a new approach to data-driven modeling. The models are flexibly configured in compliance with the neural network formalism, by accepting partially interconnected structures and various types of global and local neurons within each hidden neural layer. A simultaneous selection of convenient model structure and parameters is performed, making use of multiobjective graph genetic programming. For an efficient assessment of individuals, the authors suggest a new Pareto-ranking strategy, which permits a progressive combination between search and decision, tailored to handle objectives of different priorities. The experiments carried out for the identification of an industrial system show the capacity of the proposed approach to automatically build simple and precise models, whilst dealing with noisy data and poor aprioric information.

**Keywords**—genetic programming, multiobjective optimization, neural networks, system identification

## I. INTRODUCTION

Given a set of relevant measurements, symbolic regression can be employed to understand the behavior of any entity with which one interacts. The goal is to find the mathematical model which best illustrates the relation between the variables described by the available representative (training) data set. The usefulness of the designed model could be evaluated from different perspectives: the accuracy provided on the training data set, the interpretability, the generalization capability, etc. It should be noticed that the model is designed using a finite collection of samples acquired during a finite time interval, although its utilization must not be limited to this particular working context. Moreover, the training data set could include noisy samples, redundant information, etc., which make the generation of a model even more difficult to handle [1].

Obviously, as far as the model's adaptive capacity is concerned, it is of great interest to employ autonomic computing techniques, which permit the automatic construction of the models (keeping the transparency of any intermediary design stage), without the need of supplementary apriorical information. To this end, genetic programming (GP) offers a viable alternative, due to its efficiency in self managing a population of potential models when dealing with large search spaces, nonlinearities and discontinuities [2, 3]. It is worth mentioning that GP can simultaneously approach variable

reduction, structure selection and parameter estimation, leading to a more realistic evaluation of every generated model and increased adaptation capacity. The cost of this flexibility is an exceedingly large exploration space which can impede the search algorithm from finding a convenient solution in a reasonable amount of time. This risk can be diminished without altering the versatility of the approach by adopting a model template proven to feature universal approximation capabilities. For this, the authors considered the evolution of feed forward partially interconnected hybrid neural networks (HNN) [4, 5]. The HNN accepts any combination of hidden neurons and any connectivity map, making it possible to develop parsimonious models with increased (inductive) learning potential, fit to the particular problem to solve.

In order to be able to take advantage of the HNN flexibility, GP has to perform the search while keeping an adequate balance between the exploitation of the available genetic material (in the absence of any physical interpretation of the inner substructures of the models) and the exploration of new areas. Considering the aforementioned requirement, the paper introduces an original fitness assignment scheme able to adapt the selection pressure imposed during the evolutionary loop by multiple objectives of different priorities. The technique is applied considering that the quality of the models is assessed in terms of accuracy and parsimony. This multiobjective formulation comes naturally from application and GP standpoints, standing as an efficient anti-overfitting strategy. It should be noted that GP is configured to manage directed acyclic graphs (DAG) with the guarantee of genotypic and phenotypic validity, making use of certain extended genetic operators that are enabled to work on the functions embedded in the nodes of the graph.

The paper is organized as follows. Section II browses through similar approaches presented in literature. A brief overview of the design algorithm is given in Section III, while Section IV describes in detail the proposed fitness assignment scheme (compatible with multiobjective optimizations). Section V illustrates the applicability of the method on the identification of a nonlinear industrial system; the last section is dedicated to conclusions.

## II. RELATED WORK

A key issue in neural evolution is the encoding.

Numerous indirect and direct encryption schemes have been proposed, however most of them have been tailored to handle only a limited set of neural topologies (usually MLP or RBF) [1, 2]. Vector-based, matrix-based or tree-based individuals can lead to sparse, memory consuming representations, which are inefficiently managed when dealing with numerous neurons and connections. Although graph encoding is the most straightforward way for representing the neural models, graph GP was previously applied only for the hierarchical structures which do not assign any parameters to the neural links [6, 7], having in mind a simpler offspring production. The algorithm described in the following sections accepts various types of neurons with different number of parameters and variable arity [5].

The most natural formulation in symbolic regression is a multiobjective optimization (MOO) addressing the accuracy and parsimony of the models [1, 8]. The minimization of the individuals' complexity order provides a way to select compact models that are easier to use and have potentially better generalization capabilities. Additionally, in the case of GP, MOO acts as an efficient anti-bloat and anti-horizontal expansion technique, diminishing the risk of producing inappropriate larger structures, by means of genetic operators [2]. However, the MOO requires specific customizations of the evolutionary algorithms, as the employment of conflicting objectives leads to an infinite set of Pareto-optimal solutions, each one indicating a possible optimal compromise between accuracy and parsimony [8, 9, 10]. Several attempts have been made so far to conduct the evolutionary search toward a population of diverse solutions distributed along the Pareto-optimal set, from which the user could pick the winner (a posteriori), according to additional heuristics. When dealing with few objectives [8, 10] (as in this MOO case), Pareto techniques seem to be the most efficient ones. They set the ranks of the individuals via a dominance analysis. Higher convergence speed is ensured by the elitist Pareto methods, which gradually build an external set of elites standing as references for fitness assignment. For a successful exploration, Pareto methods need to pay special attention to preserving the diversity [8, 10]. As an attempt to fulfill this goal, the recombination could be driven by enhanced genetic operators [11]. Other approaches promote the solutions close to the knees featured by the set of nondominated solutions [12], or slightly increase the fitness of solitary dominant individuals by means of niche or crowding techniques which measure the similarity within the objective or the search space [9, 13]. Almost all MOO evolutionary methods handle objectives of equal priorities. Different levels of priorities are managed in [13], using predefined imposed targets. The fitness assignment scheme presented in Section IV adapts distinct selection pressures for the involved objectives, making use of the particularities of the MOO problem and the mean performances of the population.

### III. ALGORITHM OVERVIEW

The canonical tree-based GP has been extended to evolve a population of DAG-based individuals, each one encoding a partially interconnected HNN with global and local neurons. It should be noted that, in comparison with the homogeneous and/or the fully connected neural networks (such as MLP, RBF), these neural models can provide increased approximation capabilities, if an appropriate design of their topologies is performed [4]. The direct DAG based encoding exploits the modularity and parallelism of the neural models (a leaf node of the graph encodes a neural input, a non-leaf node corresponds to a neuron, a graph link encrypts a neural connection), leading to a compact and natural representation, which inherently ensures an efficient reuse of the model's inner substructures (e.g. a new graph link involves the implicit reuse of the corresponding sub-graph without the need of additional blocks). However, graph GP requires the extension of the canonical initialization procedure and genetic operators, which are limited to the tree-based representation.

With this in mind, the set of functions (**O**) includes the input-output mappings characterizing the types of neurons accepted in HNN layers, and the terminal set (**T**) lists all model input variables. It is worth mentioning that the static HNNs could be used for approximating dynamic functions, if **T** implements the required external delays. Assuming series - parallel schemes configured in compliance with the input-output formalism, it results that:

$$\mathbf{T} = [\mathbf{u}(k), \dots, \mathbf{u}(k - n_u), \mathbf{y}(k - 1), \dots, \mathbf{y}(k - n_y)], \quad (1)$$

where  $\mathbf{u} \in \mathbb{R}^m$  and  $\mathbf{y} \in \mathbb{R}^n$  denote the inputs and the outputs of the system to identify,  $k$  indicates the current sampling instant and  $n_u$ ,  $n_y$  represent the maximum permitted input and output lags, respectively. **T** permits to illustrate potential interdependencies between system outputs, so, without altering the universality of the approach, the HNN models could be designed as single output ones. When dealing with multiple inputs multiple outputs systems, each output should be approximated by a distinct model.

To allow a flexible connectivity between the nodes of the graph, the functions included in **O** accept variable arity and variable number of parameters. Any type of global and local neurons can be used, although this exemplification manages a combination of global perceptrons with or without functional links [14] and local Gaussian neurons with real or complex weights [15]. If  $\mathbf{z} = [z_i]_{i=1, no\_i}$  denotes the neural inputs of a neuron, the functions of **O** can be written as follows:

$$f_{PS}(\mathbf{z}, \boldsymbol{\theta}_{PS}) = \tanh\left(\sum_{i=1}^{no\_i} w_i z_i + b\right), \quad (2)$$

for the standard perceptron with weights  $w_i$  and bias  $b$

$(\theta_{PS} = [w_1 \dots w_{no-1} b]' \in \mathcal{R}^{no-i+1});$

$$f_{PF}(\mathbf{z}, \theta_{PF}) = \tanh\left(\sum_{i=1}^{no-i} \left(\sum_{j=1}^P w_{ij}^c (\cos \pi_j z_i)\right) + \sum_{j=1}^P w_{ij}^s (\sin \pi_j z_i) + w_i z_i + b\right) \quad (3)$$

for the global neuron with orthogonal trigonometric functional links of maximum order  $P$  ( $\theta_{PF} \in \mathcal{R}^{1+no-i \cdot (2P+1)}$ ) [14];

$$f_{GR}(\mathbf{z}, \theta_{GR}) = e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{no-i} (c_i - z_i)^2}, \quad (4)$$

for the Gaussian neuron with centers  $c_i$  and spread  $\sigma$

$(\theta_{GR} = [c_1, \dots, c_{no-i}, \sigma]' \in \mathcal{R}^{no-i+1});$

$$f_{GC}(\mathbf{z}, \theta_{RC}) = e^{-w^2 \left[ \left( \sum_{i=1}^{no-i} \cos \alpha_i (z_i - c_i) \right)^2 + \left( \sum_{i=1}^{no-i} \sin \alpha_i (z_i - c_i) \right)^2 \right]} \quad (5)$$

for the local neuron with centers  $c_i$  and complex weights

$w e^{\sqrt{-1} \cdot \alpha_i} (\theta_{GC} = [\alpha_1, \dots, \alpha_{no-i}, c_1, \dots, c_{no-i}, w]' \in \mathcal{R}^{2no-i+1})$  [15]. The minimally sufficient set  $\mathbf{O} = \{f_{PS}, f_{PF}, f_{GR}, f_{GC}\}$  and the sufficient terminal set (1) guarantee the phenotypic and genotypic validity, and permit the generation of any accepted HNN. Variable reduction is performed by selecting a subset of terminals within each particular model, so  $\mathbf{T}$  may include some extraneous terms.

Considering the former, the initial population of DAG based individuals is randomly generated, starting with tree-based individuals of different shapes and depths, which are recursively built by combining the elements of  $\mathbf{O}$  and  $\mathbf{T}$ . These individuals are forced to include the whole set of terminals, as basis for an efficient future variable reduction. Afterwards, extra links are added between the nodes of consecutive layers to form the initial DAG based individuals.

The offspring are produced by means of genetic operators which act on the structure and the parameters of the models: the structural crossover interchanges two sub-graphs randomly chosen from the parents; the structural mutation replaces a sub-graph with a new one, stochastically generated; the node mutation changes the type of randomly selected nodes; the link mutation adds extra-connections from the function nodes to the children of other neurons situated on the same level; the parametric mutation modifies the parameters of a selected node, without altering the type of the node or the map of the input connections. To preserve data consistency after each structural change, whilst keeping the algorithm at reasonable computational costs, the genetic operators

distinguish between normal and extra graph connections. If all extra connections are to be eliminated, a DAG would transform into a tree. Each node is connected by a normal link with only one of its parents, and by extra links with any other parents. For an adequate management of pointers and data blocks copying, any structure modification is accompanied by a correction of links' types performed for all involved nodes [5].

Unfortunately, these genetic operators can produce bloat (namely an increase of the mean complexity of the population unaccompanied by a significant improvement of the accuracy) and horizontal expansion (meaning the production of less interconnected graphs, caused by successive insertions of additional nodes due to the need of transforming some extra links to normal ones). These undesired tendencies are controlled by MOO, which rejects the exceedingly large individuals. Details are given in Section IV.

The performances of the individuals are evaluated in terms of accuracy and parsimony. Fitness assignment is solved by means of a Pareto ranking scheme which adapts the selection pressure imposed by the objectives in accordance with the mean performances of the population and the diversity of the best individuals. The resulted fitness values are used to fill the recombination pool by means of stochastic universal sampling and to control insertion (the best offspring replace the worst parents). The result of the algorithm is the most accurate model of the final population.

#### IV. ALGORITHM ENHANCEMENTS

The previously mentioned MOO features certain particularities which should be exploited by the design procedure. Firstly, some Pareto optimal solutions have no practical utilization (e.g. the solutions which are very accurate, but complex, or the solutions which are very simple, but inaccurate). Therefore, the search around these solutions could be progressively avoided. Secondly, the complexity objective is just a "proxy" one [8], introduced to prevent an undesired excessive growth of graphs. Thus, it should be assigned with a lower priority and/or exploited at a qualitative level, rather than a quantitative one. These heuristics can stand as a decision algorithm to be gradually combined with the search. With this in mind, the authors have introduced a Pareto fitness assignment procedure which enforces the search within the areas preferred by the application.

Let us consider the representative training data set,  $\mathbf{S} = \{(\mathbf{u}(t), \mathbf{y}(t))\}$ , describing the dynamic behavior of the system at instants  $t = 1, \dots, d$ . To avoid ill conditioning and inconvenient aggregation of the neurons' outputs, the samples of  $\mathbf{S}$  are scaled in  $[-0.9, 0.9]$ . The MOO aims to minimize the empirical risk (evaluated in terms of  $\mathbf{S}$ ) and the complexity order of model, expressed as indicated below:

$$f_1(M) = \frac{\sum_{t=1}^d (y_i(t) - \hat{y}_i(t))^2}{d}, \quad f_2(M) = n_{nodes} + \frac{n_{links}}{MAXL}. \quad (6)$$



Here,  $M$  denotes the evaluated model encrypted as a graph-based individual,  $y_i(t)$  and  $\hat{y}_i(t)$  denote the  $i^{\text{th}}$  output of the plant at instant  $t$  and its approximation, respectively,  $n_{\text{nodes}}$  and  $n_{\text{links}}$  specifies the number of nodes and links included in  $M$ , and  $MAXL$  represents the maximum arity of a graph node. It is worth mentioning that  $f_2$  penalizes the graphs with numerous nodes or numerous links, giving support for the diminution of bloat and horizontal expansion.

Let us assume a current population  $P(t) = \{M_i, i=1, N\}$ , with the average performances  $m_1$  and  $m_2$ , computed in terms of  $f_1$  and  $f_2$ , respectively. For a refined ranking,  $P(t)$  is firstly separated in three clusters:

$$\begin{aligned} C_1 &= \{M_j \in P(t) | f_1(M_j) \leq m_1, f_2(M_j) \leq m_2\}, \\ C_2 &= \{M_j \in P(t) | f_1(M_j) \leq m_1, f_2(M_j) > m_2\}, \quad (7) \\ C_3 &= P(t) \setminus (C_1 \cup C_2). \end{aligned}$$

and the diversity of the individuals belonging to  $C_1$  is measured by means of  $DCN$ , namely the distance within  $(f_1, f_2)$  space to the closest neighbor found in  $P(t)$ . Keeping increased diversity inside  $C_1$  is a key issue for avoiding the premature convergence of the algorithm. Therefore, if at successive generations the average  $DCN$  decreases with more than 10%, then  $m_1 \rightarrow m_1/0.8$ ,  $m_2 \rightarrow m_2/0.8$  and the clusters are redrawn in accordance with the new values  $m_{1,2}$  and (7).

Inside each resulted cluster, the Pareto-ranks are separately determined. When working on individuals who are simpler and more accurate than average (as the ones belonging to  $C_1$ ), the MOO is enforced to improve the performances of  $f_1$ , whilst keeping increased diversity. Therefore, inside  $C_1$  the Pareto-ranks are assigned in terms of  $f_1$  and  $-DCN$ , while the graphs included in  $C_2$  and  $C_3$  are separately Pareto-ranked, subject to  $(f_1, f_2)$ . Afterwards, the fitness values are linearly assigned according to the resulted ranks, keeping the selection probabilities from  $C_1$  higher than those used in  $C_2$ , and the selection probabilities from  $C_2$  higher than those used in  $C_3$ .

Whenever necessary, the Pareto-fronts are depicted in sequence, by starting with the first order one, which includes the non-dominated solutions of the set. Then, these solutions are eliminated and the second-order front is filled with the non-dominated solutions of the remaining set, and so on [9]. One considers that a solution  $M_i$  dominates a solution  $M_j$  in terms of  $f_k$  and  $f_p$ , if  $f_{k,p}(M_i) \leq f_{k,p}(M_j)$  and if  $\exists q \in \{k, p\}$ , for which  $f_q(M_i) < f_q(M_j)$ . It should be noticed that the proposed fitness assignment scheme is highly improbable

to produce fronts with numerous solutions. As the diversity of the relevant solutions is directly controlled, no additional fitness tuning is employed to recalibrate the scores of the individuals belonging to the same Pareto-front.

According to the linear ranking, the selection probability is changed with  $\frac{-s \cdot \Delta r}{N(N-1)}$  whenever the corresponding rank is increased with  $\Delta r$  ( $r=1$  corresponds to the best solutions). In order to encourage the population diversity, here  $s=0.2$ . With this in mind, one can analyze the behavior of the proposed fitness assignment scheme (CF), assuming as references the ranks given by the Pareto-fronts extracted from the whole  $P(t)$  (WF). CF and WF selections involve a comparable timeover, yet, without genetic operators, CF encourages the duplication of fewer (most accurate) solutions from the initial population. For CF, the simplest individuals of  $C_3$  and the most accurate individuals of  $C_2$  receive significantly worse ranks. The best ranks are still assigned to the individuals of  $C_1$ , so, if the genetic operators are able to produce offspring with better or similar performances than their parents,  $m_1$  and  $m_2$  decrease during the evolutionary loop, requesting the reconfiguration of clusters. Whenever  $C_1$  gains too many similar individuals via insertion, their ranks are increased in order to enforce diversity preservation. Inside  $C_1$ , even the point  $(m_1, m_2)$  can result non-dominated, when it leads to the biggest  $DCN$ . Other important variations of ranks result for the simplest individuals of  $C_1$ , which have the accuracy close to  $m_1$ . Fig. 1 illustrates these differences for a population  $P(t)$  uniformly distributed over  $(f_1, f_2)$ .

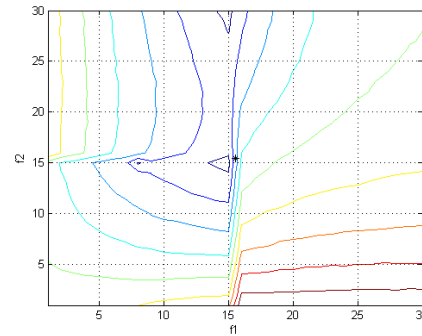


Figure 1. The contour type plot - the differences existing between the ranks given by CF and WF on a population uniformly distributed within  $(f_1, f_2)$ . "\*" marks a virtual individual with mean objective values,  $m_1$  and  $m_2$ . A line indicates a constant level. Implicit MATLAB color map is used (black-blue for the lowest levels, red for the highest ones).

Summing up, this simple fitness assignment scheme brings several important advantages: i) for the most accurate solutions, the complexity objective acts at a qualitative level, only; ii) the diversity of the best solutions is directly controlled, without additional niche or crowding techniques; iii) the preliminary clustering

combines decision and search, in accordance with the mean performances of the population, without the need of supplementary a priori information; iv) Pareto-ranking is performed in smaller clusters, at lower computational costs; v) Pareto-ranking is applied for bi-objective optimizations only, leading to increased relevance of ranking.

## V. COMPARATIVE EXPERIMENTAL VALIDATION

The applicability of the proposed algorithm (called MOCF-HGP) is demonstrated on the identification of a nonlinear industrial plant, by using measurements acquired during real-time operation. It is worth mentioning that no sufficiently accurate analytical model is accessible for describing the dependencies between the plant variables. The dynamic system is an evaporator (EV) from Lublin Sugar Factory of Poland; it admits three inputs (the steam flow and steam temperature at the input of evaporation station, as well as the juice temperature after heater) and one output (the juice temperature at the exit of the corresponding evaporation section). All missing or uncertain values are replaced with polynomial interpolation and the noise is filtered with a low - pass 4<sup>th</sup> order Butterworth system. Having in mind the generalization capability of the model, the training and the validation samples are chosen from different production shifts, showing as large as possible ranges of plant variables, for  $|S| \approx 300$ .

In order to illustrate the effects of the adopted fitness assignment procedure, the results of MOCF-HGP are compared with those provided by three other design algorithms: O-HGP (which ensures the mono-objective optimization of  $f_1$ ), MOA-HGP (which solves the MOO by means of linear aggregation, with the weights 1000 for  $f_1$  and 0.001 for  $f_2$ ) and MOWF-HGP (which assigns the ranks by means of WF). The algorithms were implemented by the authors in C/C++ and verified by using the configurations indicated in Table I. Table II shows the mean empirical risk obtained on the training and the validation data set (columns  $T$  and  $V$ , respectively) for the individuals which have been selected after five independent runs. As expected, MOO can act as an efficient anti-overfitting technique, leading to models of improved generalization capabilities (if the behavior captured by the training samples can be extrapolated to other working points). When MOO is solved by linear aggregation, the main drawback is that the weights must be apriorically set, yet improper values will impede the search, leading to unsatisfactory approximations (#1, 3, 4, 5). For sufficiently large populations, MOWF-HGP and MOCF-HGP are able to select simple models which provide significantly better approximation of the validation data set (#2, 4, 6). Unlike MOA-HGP, the final population of MOWF-HGP gives richer relevant information, as each non-dominated solution can be associated to a distinct aggregation of the objective functions. However, WF wastes part of its evolvement effort for keeping several non-dominant solutions which

are certainly unsuitable for the application. This downside is eliminated by MOCF-HGP, which makes use of additional heuristics in order to guide the search toward accurate models of reasonable complexity. The effect could be revealed when larger search spaces have to be explored (e. g., #5 and #6 - corresponding to a non-minimally sufficient  $T$ ). In these situations, MOCF-HGP is able to perform a more efficient variable reduction and to offer increased exploration speed, the result of the algorithm being introduced within the population at earlier generations. It should be noticed that real applications are confronted with such cases, whenever poor aprioric information about variable dependencies is available.

TABLE I. TABLE I. CONFIGURATIONS USED FOR EXPERIMENTAL TRIALS

#	1	2	3	4	5	6
$N$	300	1000	300	1000	300	1000
$N_{gen}$	100	300	100	300	100	300
$n_u = n_y$	1	1	3	3	5	5

$N_{gen}$  indicates the number of generations.

TABLE II. TABLE II. EXPERIMENTAL RESULTS – MOO

#	O-HGP		MOA-HGP		MOWF-HGP		MOCF-HGP	
	$T$	$V$	$T$	$V$	$T$	$V$	$T$	$V$
1	0.0035	0.0347	0.0032	0.0201	0.0025	0.0070	0.0023	0.0172
2	0.0013	0.0216	0.0021	0.0061	0.0014	0.0066	0.0012	0.0030
3	0.0025	0.0146	0.0052	0.1010	0.0029	0.0147	0.0028	0.0174
4	0.0016	0.0089	0.0024	0.0148	0.0019	0.0062	0.0016	0.0057
5	0.0025	0.0146	0.0055	0.0220	0.0035	0.0198	0.0038	0.0177
6	0.0016	0.0089	0.0033	0.0473	0.0019	0.0059	0.0017	0.0047

TABLE III. TABLE III. EXPERIMENTAL RESULTS – COMPARISON WITH MLP AND RBF

NN type	nneur	np	MSE_L_r	MSE_V_r
HNN	4	25	10.0874	11.2516
RBF	4	76	71.8516	52.0957
MLP	4	76	25.7885	32.8515

$MSE_{L_r}$  and  $MSE_{V_r}$  represent the mean squared error computed on the non-scaled learning and validation data sets;  $n_{neur}$  indicates the number of neurons and  $n_p$  specify the number of neural parameters.

An individual obtained for #4 (Table II) was selected for supplementary analysis. It ensures a better approximation (for the training and the validation data set) than an MLP or an RBF with the same number of hidden neurons (Table III and Fig.2); the MLP was trained for 3000 epochs by means of Levenberg-Marquardt algorithm, and the RBF was constructed by iteratively adding the GR neurons. It is worth mentioning that MOCF-HGP is able to take advantage of the partially interconnected HNN architectures, leading to a compact model with fewer neural parameters. The homogeneous fully connected neural networks can offer similar performances of approximation for a bigger number of hidden neurons. The residual analysis indicates insignificant relative errors (within [-0.82%, 1.29%]), 99% confidence for most lagged auto-correlation coefficients, Theil value 0.62. (approximation much better than the naïve one). Consequently, one can appreciate that this model captures the most relevant properties of the plant.

The Pareto-ranking scheme is able to preserve the diversity of the population, as indicated in Fig. 3. This means that after 300 generations the algorithm still has increased exploration capability. However, the quality of the offspring is also dependent to the productivity and the exploration strength of the genetic operators, so the achievement of better solutions during additional generations is not guaranteed. In fact, in this case, the model chosen as result of the algorithm was found at the 139<sup>th</sup> generation.

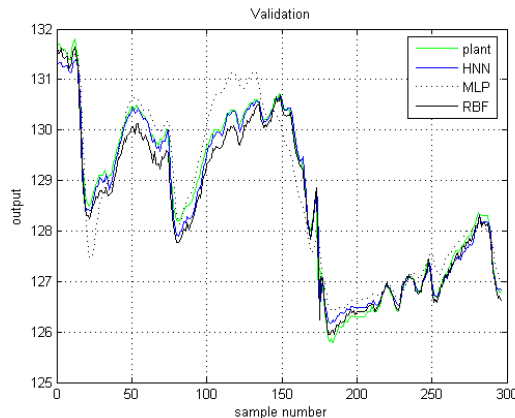


Figure 2. The approximation of the validation data set, provided by HNN, MLP and RBF.

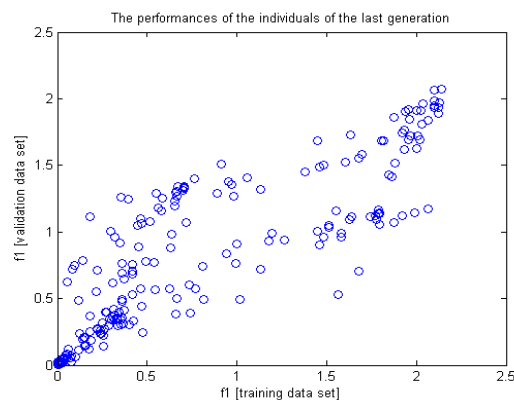


Figure 3. The accuracy performances provided by the individuals of the final population.

## VI. CONCLUSIONS

The proposed multiobjective GP-based algorithm allows an efficient design of accurate and compact HNNs, within the framework of data driven modeling. Every generation, the individuals are separated in three distinct clusters, each one with a specific ranking scheme. The suggested procedure directly controls the diversity of the most accurate solutions and reshapes the clusters in accordance with the average performances of the population, lying in improved exploration capabilities.

As the complexity objective discards the over-fitted models constructed during the evolutionary loop, MOO graph GP can deal with a non-minimally sufficient terminal set and provides increased efficiency in avoiding bloat and horizontal expansion.

The approach is suitable for the design of dynamic or static nonlinear models, whenever the interdependency between the involved variables is not completely understood, large search spaces have to be scanned, or/and high model accuracy is required.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support offered for the completion of this work by "SICONA" research grant 12100/2008, which is sustained by The National Centre for Programs Management (Romania).

## REFERENCES

- [1] P. J. Flemming, R. C. Purshouse "Evolutionary Algorithms in Control Systems Engineering: A Survey", in *Control Engineering Practice* 10, 2002, pp. 1223-1241.
- [2] Poli R., Langdon W. B., Mc Phee N. F., *A Field Guide to Genetic Programming*. Published via <http://lulu.com> (with contributions of J.R. Koza), [Online]. Available: <http://www.gp-field-guide.org.uk>, 2008.
- [3] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge, MA: MIT Press, 1992.
- [4] L. Ferariu, M. Voicu, "Nonlinear System Identification Based on Evolutionary Dynamic Neural Networks with Hybrid Structure" in *Proc. of IFAC Congress, Prague, Czech Republic*, 2005.
- [5] L. Ferariu, B. Burlacu B., "Evolutionary Neural Networks with Heterogeneous Hidden Layers", in *Transactions on Automatic Control and Computer Science, Scientific Bulletin of "Politehnica" University of Timisoara*, 55 (69), 4, 2010, pp. 239-246.
- [6] A. J. Barton, J. J. Valdés, R. Orchard, "Neural networks with multiple general neuron models: A hybrid computational intelligence approach using Genetic Programming", in *Neural Networks*, 22, 2009, pp. 614-622.
- [7] A. Walker, J. F. Miller, "The Automatic Acquisition, Evolution and Reuse of Modules in Cartesian Genetic Programming", in *IEEE Transactions on Evolutionary Computation*, 12 (4), 2008, pp. 397-417.
- [8] J. Handl, J. Knowles, "Modes of Problem Solving with Multiple Objectives: Implications for Interpreting the Pareto Set and for Decision Making" in *Multiobjective Problem Solving from Nature*, J. Knowles · D. Corne, · K. Deb (Eds.), New York: Springer, 2008, pp.131-154.
- [9] K. Deb, *Multiobjective Optimization Using Evolutionary Algorithms*, Wiley&Sons, 2001.
- [10] E. J. Hughes, "Fitness Assignment Methods for Many-Objective Problems", in *Multiobjective Problem Solving from Nature*, J. Knowles · D. Corne, · K. Deb (Eds.), New York: Springer, 2008, 307-330.
- [11] G. Chen, C. P. Low, "Preserving and Exploiting Genetic Diversity in Evolutionary Programming Algorithms", in *IEEE Transactions on Evolutionary Computation*, 13 (3), 2009, pp. 661-673.
- [12] L. Rachmawati, D. Srinivasan, "Multiobjective Evolutionary Algorithm with Controllable Focus on the Knees of the Pareto Front" in *IEEE Transactions on Evolutionary Computation*, 13 (4), 2004, pp. 810-824.
- [13] K. Rodriguez-Vasquez., C. M Fonseca, P. J. Fleming, "Identifying the Structure of Nonlinear Dynamic Systems Using Multiobjective Genetic Programming" in *IEEE Transactions on Systems Man and Cybernetics, Part A – Systems and Humans*, 34, 2004, pp. 531-534.
- [14] J. Patra, R. Pal, B. Chatterji, G. Panda, "Identification of nonlinear dynamic systems using functional link artificial neural networks", in *IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics*, 29, 1999, pp. 254-262.
- [15] B. Igennik, M. Tabib – Azar, Le Clair S. R., "A net with complex weights", in *IEEE Transactions on Neural Networks*, 12, 2001, pp. 236-249.

# A regressive schema theory based tool for GP evolved nonlinear models

Alina Patelli, Lavinia Ferariu  
Dept. of Automatic Control and Applied Informatics  
“Gh. Asachi” Technical University of Iasi  
Iasi, Romania  
{apatelli, lferaru}@tuiasi.ro

**Abstract**—Nonlinear systems identification is approached by employing a genetic programming computational tool featuring explicit building block exploitation. The level of adaptation of recurrent model sub-structures is assessed by a fuzzy module. The first contribution of the paper resides in using the fuzzy classification results to reconfigure the cut point selection probabilities of regressor inner nodes, a process called encapsulation. This allows for the second innovation, namely the design of context aware genetic operators capable of protecting the existing instances of fit building blocks and of creating new ones. The computational costs of encapsulation are reduced by employing a novel regressive schema theory – the third and main paper contribution – which assesses the inherent chances of regressor survival. A thorough theoretical support for demonstrating the efficiency of context aware operators in transmitting schema instances over the generations is introduced. The suggested algorithm is experimentally validated in the framework of a complex, industrial, nonlinear subsystem of a sugar factory.

**Keywords**—multiobjective optimisation, nonlinear models, fuzzy logics, genetic programming

## I. INTRODUCTION

Most realistic identification problems involve multiple conflicting optimization objectives [1]. Solving such a problem implies generating a set of nondominated solutions, conveniently distributed over specific areas of the Pareto front, which correspond to the desired objectives trade-offs [2]. One of the keynote advantages that evolutionary approaches (EAs) have over classical identification techniques is the capacity to produce several nondominated solutions in a single run [3]. By evolving an entire population of candidate models at a time, EAs are, in general, more robust than deterministic methods [4]. When dealing with model structure selection, traditional techniques offer limited support (e.g. NARMAX polynomial models may be built by sequentially adding or pruning terms), whilst EAs are capable of extensively exploring the search space, requiring a reduced amount of *a priori* information [2],[3].

In this context, genetic programming (GP) features the additional capability of encrypting nonlinear models in a hierarchical fashion, inherently compatible with computational tools [4]. A consequence of evolving hierarchical recursive structures is the possibility of explicitly exploiting building blocks (BBs), which represent model sub-components. This way, decision

making may be implemented at a microscopic level, by encouraging the survival of certain features exhibited by fit individuals (situated in the feasible areas of the Pareto front). Macroscopic exploitation promotes fit individuals, as a whole, which may generate a limited number of children at a certain generation. In that respect, the microscopic approach is more effective, as a well adapted building block may be transmitted to a theoretically unrestricted number of children from the following generation.

A convenient way of building hierarchical models is to consider the nonlinear, linear in parameters (NLP) formalism, as it is a universal approximator, capable of encrypting any bounded function to any degree of accuracy [5]. NLP compliant models are linear combinations (a sum) of nonlinear terms (products) called regressors, thus it requires only two operators to connect tree nodes, namely “+” and “\*”. The presented approach considers a fuzzy module (FM) to classify regressors according to their level of adaptation (i.e. the extent to which regressors influence their host individual’s accuracy and parsimony). An innovative encapsulation mechanism is also suggested, in order to protect fit BBs from the potentially disruptive action of the genetic operators.

In order to reduce the computational resource consumption implied by applying encapsulation, the authors introduce an original regressive schema theory (RST) used to evaluate the inherent chances of BB survival. This way, only regressors with a slim intrinsic transmission probability are protected via encapsulation from the potentially disruptive effect of genetic operators. The RST is a one step ahead prediction instrument which only makes use of information at the current generation to compute the expected number of regressor instances at the following iteration. As shown later in this paper, RST may also be used to train context aware genetic operators to protect the integrity of existent fit BB instances as well as create new ones. By providing crisp results, as opposed to lower or upper bounds, RST offers support for a rigorous theoretical demonstration of encapsulation efficiency. Considering the amount of insight that it provides, relative to the behavior of the GP algorithm, the RST is capable of generating a model for the process of evolutionary search in itself.

This paper introduces a novel RST based method to compute regressor transmission probabilities whilst considering a set of genetic operators comprised of cut

point crossover and point mutation. Secondly, the efficiency of the encapsulation procedure in determining the genetic operators to protect existing schema instances and create new ones is mathematically demonstrated. Thirdly, a general formula for the expected number of regressor instances at the following generation is also provided.

The following section overviews the recent state of the art in the field of nonlinear systems identification by evolutionary means, as well as in the domain of schema theories. Section 3 offers details about the logical predicates necessary to define the transmission probabilities described in section 4. The concept of encapsulation, as well as the mathematical proof of its efficiency, is the topic of section 5. Section 6 contains an experimental validation of the suggested algorithm to confirm the theoretical results previously presented, whilst the final section is dedicated to conclusions.

## II. STATE OF THE ART

Previous attempts of employing GP for nonlinear systems identification range from monoobjective approaches [6] to multiobjective methods [7], [8] and memetic techniques [6], [9]. Madar *et al.* suggest evaluating regressive models solely according to their approximation accuracy, and employing orthogonally least squares methods for parameters computation [6]. The identification problem is approached from a multiobjective standpoint in [7], where an external preference vector is used to dictate the weights of the considered optimization criteria. In the context of symbolic regression, Vladislavleva proposes a new indicator of a model's generalization capacity, namely the order of nonlinearity [8]. It is used to evaluate the smoothness of the model's response over the validation data set, thus assessing its extrapolation capability. In order to dynamically configure objectives' priorities, [9] suggests an elite clustering technique as well as a double threaded offspring generation process over the elite population and the regular one.

All the above mentioned techniques employ an implicit BB exploitation scheme as they encourage the propagation of fit individuals, which are highly likely to feature well adapted BBs. The algorithm suggested in [10] explicitly exploits structural blocks by considering an initial population of all BBs of a certain size, that can be built with the available primitives. The blocks are combined into coherent models, in the juxtapositional phase, by means of cut and splice recombination operators. In [11], a method is provided to trace BBs throughout the search process. It is capable to gradually determine the optimal block size in order to solve a certain problem, as well as the stability of the involved variables.

According to [12], the population evolved by an EA presents itself as a cloud of points, whose shape and trajectory throughout the search space are governed by stochastic rules. Increasing GP efficiency for specific applications is possible without determining the exact coordinates of the cloud at any time [3], [8], [9]. However, such efforts would become straightforward if a model of the cloud's shape and trajectory were available [12]. One

way to attain that goal is to capture the dynamics of recurrent structural patterns (BBs) in the population, by means of a schema theory. First acknowledged in [4], the schema concept has seen a series of definitions, one of the most comprehensive being the one suggested by Poli and McPhee [12]. The two authors state that a schema is a tree that contains both primitives and wildcard symbols. The latter can be replaced with any valid primitive. Their paper also introduces the first documented exact schema theory for fixed sizes and shapes, used to calculate the expected number of schema instances at the following generation.

The schema theory suggested in [12] is valid under proportionate selection and assumes the exclusive use of the crossover operator. An extension for both crossover and mutation is available in [13], and it refers only to linear chromosomes. The present paper suggests a novel alternative to the case of fixed sizes and shapes, namely regressive schemata, capable of handling various dimensions and structures. An early version of the variable size and shape schema theory was suggested in [14], where it is used merely to monitor the efficiency of regressor encapsulation. The current variant is designed as a local search procedure over the space of all possible offspring. It is meant to train both cut point crossover and mutation to select the appropriate nodes within the parent trees so that the transmission of fit regressors is encouraged.

## III. SCHEMATA CREATION AND DESTRUCTION

Given the recursive nature of hierarchical structures, the author proposes the notation  $R_{i,k}^T$  to allow a flexible definition of a regressor, relative to any of its inner nodes  $i$ . Thus, in Figure 1., the subtree rooted in node 4 of  $T_1$  can be defined in relation to its root  $R_{4,0}^{T_1}$  or to any of its other nodes,  $R_{5,1}^{T_1} = R_{6,1}^{T_1} = R_{7,2}^{T_1} = R_{8,2}^{T_1}$ . As shown in the previous example,  $k$  represents the number of tree levels between the regressor root and node  $i$ .

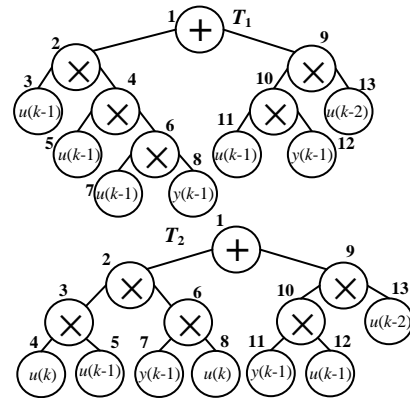


Figure 1. Trees featuring various types of regressors

The “\*” operator is commutative, thus one regressor may have various structural forms (e.g.  $R_{9,0}^{T_1}$  and  $R_{9,0}^{T_2}$  are mathematically identical). The variable size and shape schema theory hereby proposed is capable of detecting all structural instances of a given regressor, which is compulsory for the correct computation of transmission

probabilities. To enable this feature, the authors suggest the use of extended sets.

An extended set describes a regressor by enumerating its terminals and their frequency (the number of times they appear within the regressor structure), without including any information regarding node position. Thus, the extended set associated to  $R_{9,0}^{T_1}$  is  $S(R_{9,0}^{T_1}) = \{(u(k-1),1), (u(k-2),1), (y(k-1),1)\}$ , which is also a match to  $R_{9,0}^{T_2}$ . In the context of RST, a schema refers to an extended set that can be matched to all structural instances of a certain regressor.

A schema  $H$  is destroyed (cut) by the crossover operator if the selected cut point,  $i$ , is an inner node of  $H$  (selecting nodes 5, 6, 7 or 8 would lead to the disruption of  $R_{4,0}^{T_1}$ ). Mutation can damage a regressor in a similar way. The destruction of a regressor, by either genetic operator, validates predicate  $\delta_{cut}(i)$ :

$$\delta_{cut}(i) = \sum_k \delta(S(H) = S(R_{i,k}^T)), k \geq 1. \quad (1)$$

The crossover operator is capable of splicing a new instance of  $H$ , if the remaining part of the direct parent, after the removal of the swap subtree, is joined with an appropriate fragment of the indirect parent. This corresponds to the situation where  $\delta_{join}(i, j)$  is true:

$$\delta_{join}(i, j) = \sum_k \delta(S(H) = S(R_{i,k}^{T_1}) \setminus S(R_{i,0}^{T_1}) \cup S(R_{j,0}^{T_2})), k \geq 1. \quad (2)$$

In (2),  $i$  and  $j$  represent the cut points selected inside parents  $T_1$  and  $T_2$  respectively, and  $\delta$  returns the truth value of its input expression.

When terminal wise mutation is considered, the creation of a new schema instance is possible if the direct parent features a “flawed”  $H$  structure, which differs from the correct one by one terminal only. If that specific terminal,  $w_i$ , is selected by mutation, and replaced with the appropriate element,  $x_j$ , of the terminal set, then the instance of  $H$  is “repaired”, and the  $\delta_{repair}(i, j)$  predicate is true:

$$\delta_{repair}(i, j) = \sum_k \delta(S(H) = S(R_{i,k}^{T_1}) \setminus (w_i, 1) \cup (x_j, 1)), k \geq 1. \quad (3)$$

#### IV. SCHEMATA TRANSMISSION PROBABILITIES

After a sufficient number of algorithm iterations, the trees on the first order Pareto front are most likely to feature fit regressors. Therefore, a common regressor search (CRS) procedure identifies the BBs with multiple instances inside the structure of the nondominated trees. The degree of adaptation of each regressor located by the CRS is estimated by a fuzzy module (FM), which assigns every BB with a discrete valued adaptation estimation label (AEL). Details about the inner logic of the fuzzy classifier and the configuration of its parameters are available in [9].

The regressors with an  $AEL \geq 0.8$  are most likely fit structural blocks. If transmitted in multiple copies to the next generation, such BBs would enhance the quality of

their host individuals. Some of them will survive inherently, while others will need supplementary protection from the effect of the genetic operators. In order to determine which BB requires shielding against crossover and mutation, the intrinsic transmission probabilities of all regressors identified by the CRS are computed as follows.

Considering a reproduction pool of  $N$  parents organized in  $N/2$  pairs, each pair of resulting children may feature one less  $H$  instance, as many  $H$  instances, or one extra  $H$  instance, relative to their two parents. Please note that any  $H$  schema instance completely eliminated from one parent, will be transferred to the other, therefore RST only considers the destruction/creation of schema instances. Given the cut point crossover and terminal wise mutation genetic operators, the probabilities associated to the three possible scenarios mentioned above are:

$$\alpha^a(H, t) = p_{co} \alpha_{co}^a(H, t) + p_{mt} \alpha_{mt}^a(H, t), a = -1, 0, +1. \quad (4)$$

In (4),  $p_{co}$  and  $p_{mt}$  represent the probabilities of applying the crossover and the mutation operators, respectively. The probabilities of decreasing –  $\alpha_{co}^{-1}(H, t)$ , preserving –  $\alpha_{co}^0(H, t)$  or increasing –  $\alpha_{co}^{+1}(H, t)$  the number of  $H$  instances featured by the two parents, when applying the crossover operator, are:

$$\begin{aligned} \alpha_{co}^{-1}(H, t) &= \sum_{\substack{i \in T_1 \setminus \{+\} \\ j \in T_2 \setminus \{+\}}} p(i | T_1, t) p(j | T_2, t) \delta_{cut}(i) \overline{\delta_{join}(i, j)} \\ \alpha_{co}^0(H, t) &= \sum_{\substack{i \in T_1 \setminus \{+\} \\ j \in T_2 \setminus \{+\}}} p(i | T_1, t) p(j | T_2, t) (\delta_{cut}(i) \overline{\delta_{join}(i, j)} + \delta_{cut}(i) \delta_{join}(i, j)) \\ \alpha_{co}^{+1}(H, t) &= \sum_{\substack{i \in T_1 \setminus \{+\} \\ j \in T_2 \setminus \{+\}}} p(i | T_1, t) p(j | T_2, t) \delta_{cut}(i) \delta_{join}(i, j), \end{aligned} \quad (5)$$

where  $\delta_{cut}(i)$  and  $\delta_{join}(i, j)$  are defined in (1) and (2) respectively, and  $t$  stands for the current generation.

The creation, preservation and destruction probabilities, in the case of applying the terminal wise mutation operator may be analogously defined as follows:

$$\begin{aligned} \alpha_{mt}^{-1}(H, t) &= \sum_{\substack{i \in T \setminus \{+\} \\ x_j \in \mathbf{x}}} p(i | T, t) p(x_j | \mathbf{x}, t) \delta_{cut}(i) \overline{\delta_{repair}(i, j)} \\ \alpha_{mt}^0(H, t) &= \sum_{\substack{i \in T \setminus \{+\} \\ x_j \in \mathbf{x}}} p(i | T, t) p(x_j | \mathbf{x}, t) (\delta_{cut}(i) \overline{\delta_{repair}(i, j)} + \delta_{cut}(i) \delta_{repair}(i, j)) \\ \alpha_{mt}^{+1}(H, t) &= \sum_{\substack{i \in T \setminus \{+\} \\ x_j \in \mathbf{x}}} p(i | T, t) p(x_j | \mathbf{x}, t) \delta_{cut}(i) \delta_{repair}(i, j), \end{aligned} \quad (6)$$

where  $\delta_{cut}(i)$  and  $\delta_{repair}(i, j)$  are defined in (1) and (3) respectively. The results presented in (5) and (6) are generalized by the following theorems.

**Theorem 1:** Let  $T_1$  and  $T_2$  be the parents of children  $C_1$  and  $C_2$ , and let  $h$  be the number of  $H$  schema instances featured by  $T_1$  and  $T_2$ . The probabilities that  $C_1$  and  $C_2$  feature  $h-1$ ,  $h$  or  $h+1$  instances of schema  $H$ , if only crossover is considered, are:

$$\alpha_{co}^a(H, t) = \sum_{T_1, T_2} p(T_1, t) p(T_2, t) \sum_{\substack{i \in T_1 \setminus \{+\} \\ j \in T_2 \setminus \{+\}}} p(i | T_1, t) p(j | T_2, t) \delta_a(i, j) \quad (7)$$



$a \in \{-1, 0, +1\}$ ,  $\delta_{+1}(i, j) = \overline{\delta_{cut}(i)\delta_{join}(i)}$ ,  $\delta_{-1}(i, j) = \overline{\delta_{join}(i)\delta_{cut}(i)}$ , and  $\delta_0(i, j) = (\overline{\delta_{cut}(i)\delta_{join}(i, j)} + \delta_{cut}(i)\delta_{join}(i, j))$ .

#### Demonstration:

Random variable  $\delta_a(i, j)$  is Bernoulli, therefore it assumes value 1 with a probability of  $\alpha^a(H, t)$  and value 0 with a probability of  $1 - \alpha^a(H, t)$ . The expected value of  $\delta_a(i, j)$  is  $E[\delta_a(i, j)] = \sum_{T_1, T_2, i, j} \delta_a(i, j) p(i, j) = \alpha^a(H, t)$ . As

$p(i, j) = p(i|T_1, t)p(j|T_2, t)p(T_1, t)p(T_2, t)$ , by substituting this result into the previous one, (7) is obtained.

**Theorem 2:** Let  $T$  be an individual produced by crossover, let  $C$  be the offspring produced after mutating  $T$  and let  $h$  be the number of  $H$  schema instances featured by  $T$ . The probabilities that  $C$  features  $h-1$ ,  $h$  or  $h+1$  instances of schema  $H$ , are:

$$\alpha_{mi}^a(H, t) = \sum_T p(T, t) \sum_{\substack{i \in T \setminus \{+\} \\ x_j \in \mathbf{x}}} p(i|T, t)p(x_j|\mathbf{x}, t)\delta_a(i, j) \quad (8)$$

$a \in \{-1, 0, +1\}$ ,  $\delta_{+1}(i, j) = \overline{\delta_{cut}(i)\delta_{repair}(i)}$ ,  $\delta_{-1}(i, j) = \overline{\delta_{repair}(i)\delta_{cut}(i)}$ , and  $\delta_0(i, j) = (\overline{\delta_{cut}(i)\delta_{repair}(i, j)} + \delta_{cut}(i)\delta_{repair}(i, j))$ .

The demonstration is analogous to the one provided for Theorem 1.

The total transmission probability of a schema, when both crossover and mutation are applied, can be evaluated by substituting (7) and (8) in (4).

#### V. INCREASING BUILDING BLOCK TRANSMISSION PROBABILITY VIA ENCAPSULATION

The existing instances of structural blocks with  $\alpha^{-1}(H, t) < 0.2$  are unlikely to be destroyed. It is highly probable that new instances of BBs with  $\alpha^{+1}(H, t) > 0.8$  will be created. If both conditions mentioned above are met by a certain regressor, then its inherent chances of being transmitted into the next generation are satisfactory. Such BBs need not be shielded from the genetic operators, as that would imply wasting computational resources. All other regressors,  $\alpha^{-1}(H, t) \geq 0.2$  and  $\alpha^{+1}(H, t) \leq 0.8$ , require explicit protection against crossover and mutation in order to encourage their propagation.

Cut point crossover selects two nodes,  $i$  and  $j$ , one in each of the two involved parents, with a certain probability,  $p(i|T, t)$  and  $p(j|T, t)$ . The subtrees rooted in  $i$  and  $j$  are afterwards swapped, resulting in two offspring individuals. Point mutation only processes one parent at a time and selects one terminal node with a probability of  $p(q|T, t)$ . The selected node is then replaced with a valid terminal. The default values of the three selection probabilities are uniformly distributed,  $p(i|T, t) = p(j|T, t) = 1/n$ , where  $n$  is the number of nodes eligible for crossover ("\*" and terminals), whilst  $p(q|T, t) = 1/z$ , where  $z$  is the number of terminals in  $T$ .

Encapsulation (not to be mistaken for the genetic operator introduced in [4]) is the process of lowering the three selection probabilities,  $p(i|T, t)$  and  $p(j|T, t)$  in the case of crossover and  $p(q|T, t)$  relative to mutation, for all inner nodes of a regressor with slight inherent chances of survival ( $\alpha^{-1}(H, t) \geq 0.2$  and  $\alpha^{+1}(H, t) \leq 0.8$ ). The probability reduction is performed proportionally to the regressor's AEL. Before applying crossover, encapsulation consists in applying the following formulae:

$$p_{enc}(i|T) = \frac{1 - AEL(R_{i,k}^T)}{NN_{\max}}, i \in R_{i,k}^T, AEL(R_{i,k}^T) \geq 0.8 \quad (9)$$

$$p_{enc}(i|T) = \frac{1 - \sum_{R_{j,0}^T, AEL(R_{j,0}^T) \geq 0.8} RS(R_{j,0}^T) \cdot p_{enc}(l \in R_{j,0}^T | T)}{NN_{<0.8}}, AEL(R_{i,k}^T) < 0.8$$

In (9),  $RS(R_{j,0}^T)$  returns the size of the input regressor,  $NN_{\max}$  denotes the number of nodes featured by the largest encapsulated regressor in the parent population and  $NN_{<0.8}$  is the number of eligible inner nodes of non-encapsulated regressors within tree  $T$ .

Before applying mutation, encapsulation is repeated as this genetic operator only targets terminal nodes. The encapsulation formulae are:

$$p_{enc}(i|T) = \frac{1 - AEL(R_{i,k}^T)}{NZ_{\max}}, i \in R_{i,k}^T, AEL(R_{i,k}^T) \geq 0.8 \quad (10)$$

$$p_{enc}(i|T) = \frac{1 - \sum_{R_{j,0}^T, AEL(R_{j,0}^T) \geq 0.8} TS(R_{j,0}^T) \cdot p_{enc}(w \in R_{j,0}^T | T)}{NZ_{<0.8}}, AEL(R_{i,k}^T) < 0.8$$

where  $NZ_{\max}$  represents the maximum number of terminals featured by one specific regressor in the parent population,  $NZ_{<0.8}$  stands for the number of terminals in the non-encapsulated regressors, and  $TS(R_{j,0}^T)$  returns the number of terminals featured by its input regressor. Please note that the encapsulation process affects all eligible nodes inside a tree, namely the ones inside regressors with  $AEL \geq 0.8$ , as well as the ones belonging to BBs with  $AEL < 0.8$ , so that the total sum remains 1.

After encapsulation, the transmission probabilities are recomputed considering the new node selection probabilities. Considering a given pair of parents,  $T_1$  and  $T_2$ , and by substituting (9) in (5), the transmission probabilities under crossover encapsulation are obtained:

$$\alpha_{co,enc}^{-1}(H, t) = \frac{n(1 - AEL(H))}{NN_{\max}} \alpha_{co}^{-1}(H, t) \quad (11)$$

$$\alpha_{co,enc}^{+1}(H, t) = \frac{n(1 - \sum_q h_q \frac{1 - AEL(H)}{NN_{\max}})}{n - \sum_q h_q} \alpha_{co}^{+1}(H, t)$$

where  $n$  is the number of nodes eligible for crossover in  $T_1$  and  $h_q$  is the number of nodes in the  $q^{th}$  encapsulated regressor of  $T_1$ . Considering the mutation operator, the

transmission probabilities under encapsulation are obtained by substituting (10) in (6):

$$\alpha_{mt,enc}^{-1}(H,t) = \frac{z(1-AEL(H))}{NZ_{\max}} \alpha_{mt}^{-1}(H,t) \quad (12)$$

$$\alpha_{mt,enc}^{+1}(H,t) = \frac{z(1-\sum_q w_q \frac{1-AEL(H)}{NZ_{\max}})}{z-\sum_q w_q} \alpha_{mt}^{+1}(H,t),$$

where  $z$  is the number of terminals in  $T_1$  and  $w_q$  is the number of terminals in the  $q^{th}$  encapsulated regressor of  $T_1$ . Please note that theorems 1 and 2 also apply for the transmission probabilities under encapsulation, namely (11) and (12) respectively.

Encapsulation allows the transmission of more schema instances to the following generation than if default selection probabilities are considered ( $p(i|T,t)=p(j|T,t)=1/n$  and  $p(q|T,t)=1/z$ ). To demonstrate that affirmation, let us consider parents  $T_1$  and  $T_2$  and compute the expected number of regressors matching schema  $H$ , featured by the two children solutions:

$$E[F(H,t+1)] = \alpha^{-1}(H,t)(F(H,t)-1) + \alpha^0(H,t)F(H,t) + \alpha^{+1}(H,t)(F(H,t)+1) \quad (13)$$

In (13),  $\alpha^{-1}(H,t)$ ,  $\alpha^0(H,t)$ ,  $\alpha^{+1}(H,t)$  are given by (4),  $F(H,t+1)$  is the number of  $H$  occurrences ( $H$  frequency) within the two generated children at generation  $t+1$ , and  $F(H,t)$  represents the number of  $H$  schemata featured by the two parents at generation  $t$ . If encapsulation is considered, (13) becomes:

$$E[F_{enc}(H,t+1)] = \alpha_{enc}^{-1}(H,t)(F(H,t)-1) + \alpha_{enc}^0(H,t)F(H,t) + \alpha_{enc}^{+1}(H,t)(F(H,t)+1) \quad (14)$$

where  $\alpha_{enc}^{-1}(H,t)$ ,  $\alpha_{enc}^0(H,t)$ ,  $\alpha_{enc}^{+1}(H,t)$  can be obtained from (4) by adding the *enc* subscript to every  $\alpha$  probability term.

By subtracting (13) from (14), the following results:

$$\alpha_{enc}^{+1}(H,t) - \alpha^{+1}(H,t) - (\alpha_{enc}^{-1}(H,t) - \alpha^{-1}(H,t)) = \quad (15)$$

$$p_{co}(\alpha_{co,enc}^{+1}(H,t) - \alpha_{co}^{+1}(H,t) - \alpha_{co,enc}^{-1}(H,t) + \alpha_{co}^{-1}(H,t)) +$$

$$p_{mt}(\alpha_{mt,enc}^{+1}(H,t) - \alpha_{mt}^{+1}(H,t) - \alpha_{mt,enc}^{-1}(H,t) + \alpha_{mt}^{-1}(H,t))$$

After making all necessary substitutions, the sign of the first term of the right hand side sum in equation (15) can be determined by analyzing the following expression:

$$D_{co} = \alpha_{co}^{+1}(H,t) \frac{\sum_q h_q (1 - \frac{n}{NN_{\max}}) + \frac{n}{NN_{\max}} \sum_q h_q AEL(H_q)}{n - \sum_q h_q} +$$

$$\alpha_{co}^{-1}(H,t) \frac{NN_{\max} - n + nAEL(H)}{NN_{\max}} \quad (16)$$

The number of nodes in  $T_1$  is lower than in the largest individual in the population,  $NN_{\max} \geq n$ . The FM output,  $AEL(H)$  is a positive value, whereas the cumulated size of all encapsulated regressors in  $T_1$  is smaller than the total number of  $T_1$  nodes,  $\sum_q h_q \leq n$ . Therefore (16) is positive.

The second term of the right hand side sum in equation (15) can be written as:

$$D_{mt} = \alpha_{mt}^{+1}(H,t) \frac{\sum_q w_q (1 - \frac{z}{NZ_{\max}}) + \frac{z}{NZ_{\max}} \sum_q w_q AEL(H_q)}{n - \sum_q h_q} +$$

$$\alpha_{mt}^{-1}(H,t) \frac{NZ_{\max} - z + nAEL(H)}{NZ_{\max}} \quad (17)$$

Given a similar reasoning as the one before, the quantity in (17) is also positive. Thus the difference in (15) is positive, demonstrating that, in the presence of encapsulation, the resulting two children are expected to feature more schema instances than their parents.

Relative to the entire set of resulting children, the difference of expected values in (15) may be computed as follows:

$$E[F(H,t+1)] - E[F_{enc}(H,t+1)] = \sum_{T_1, T_2} p_{co} D_{co} + p_{mt} D_{mt} \geq 0 \quad (18)$$

## VI. APPLICATION

To confirm (18), the fuzzy controlled encapsulation – schema theory for regressive models – memetic evolutionary algorithm (FCE-STRM-MEA III) was deployed to identify a complex nonlinear system, namely the steam subsection of a sugar factory in Poland. The results in TABLE I. were obtained by evolving a population of 50 individuals, over 70 generations. The FM was first deployed at generation 30.

TABLE I. ENCAPSULATED REGRESSORS ANALYSIS

Nr.	AEL	Common regressors $H$	$\alpha^{+1}(H,t)$	$\alpha^{-1}(H,t)$	$F_H$
1	0.8	$u(k)y(k-1)$	0.453	0.411	3
2		$y(k-2)u(k)$	0.471	0.174	4
3		$u(k)u(k-1)u(k-2)$	0.711	0.204	2
4		$u(k-2)y(k-1)u(k)$	0.813	0.107	5
5	0.9	$y(k-2)y(k-1)$	0.803	0.112	3
6		$y(k)u(k-2)$	0.012	0.612	2
7	1	$u(k-1)u(k)y(k-2)$	0.211	0.203	2
8		$u(k-2)u(k)$	0.523	0.231	3

a. Situation at generation 30

Nr.	AEL	Common regressors $H$	$\alpha^{+1}(H,t)$	$\alpha^{-1}(H,t)$	$F_H$
1	0.8	$u(k)y(k-1)$	0.401	0.271	4
2		$y(k-2)u(k)$	0.423	0.114	4
3		$u(k)u(k-1)u(k-2)$	0.211	0.762	2
4		$u(k-2)y(k-1)u(k)$	0.123	0.614	4
5	0.9	$y(k-2)y(k-1)$	0.781	0.012	4
6		$y(k)u(k-2)$	0.213	0.312	2
7	1	$u(k-2)u(k)$	0.254	0.618	2

b. Transmission probabilities without encapsulation – gen 31

Nr.	AEL	Common regressors $H$	$\alpha_{enc}^{+1}(H,t)$	$\alpha_{enc}^{-1}(H,t)$	$F_H$
1	0.8	$u(k)y(k-1)$	0.234	0.544	6
2		$u(k)u(k-1)u(k-2)$	0.632	0.231	5
3	0.9	$y(k-2)u(k)$	0.645	0.023	7
4		$y(k)u(k)u(k)$	0.543	0.453	5
5		$y(k-2)y(k-1)$	0.432	0.132	5
6		$y(k)u(k-2)$	0.372	0.321	8
7	1	$u(k-2)y(k-2)$	0.532	0.221	4
8		$u(k-2)y(k-1)u(k)$	0.321	0.625	6
9		$u(k-1)u(k)y(k-2)$	0.376	0.154	6
10		$u(k-2)u(k)$	0.515	0.123	7
11		$y(k-2)y(k-2)y(k-1)$	0.324	0.514	3

c. Transmission probabilities under encapsulation – gen 31

The behavior of a basic algorithm, employing no encapsulation, is described in TABLE I. As no regressor is protected, entry 7 in TABLE I. is no longer identified by the CRS at generation 31 (there is only one regressor with  $AEL = 1$  in TABLE I. ), which shows that at least one of its two instances was destroyed by crossover. The second regressor with  $AEL = 1$ , entry 8 in TABLE I. , transmits two instances in generation 30, yet still loses one, an undesired behavior, given the fact that its fuzzy degree of adaptation is maximum. Also note that no new regressors with  $AEL \geq 0.8$  are created.

By running FCE-STRM-MEA III, a noticeable increase can be observed in the number of regressor instances, for all considered  $AELs$  (TABLE I. ), as anticipated by the result in (18). Entry 4 in TABLE I. jumps from  $AEL = 0.8$  at generation 30 to  $AEL = 1$  at generation 31, which is most likely a sign of increased exploitation. Entry 2 in TABLE I. also jumps from  $AEL = 0.8$  to  $AEL = 0.9$ . The exploratory power of FCE-STRM-MEA III is illustrated by the discovery/creation of five new fit regressor instances, namely entries 4,5,7,8,11 in TABLE I. .

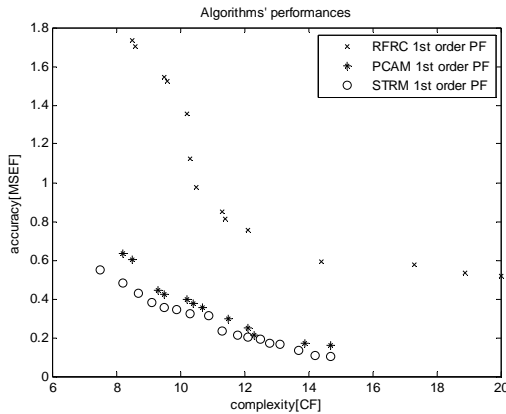


Figure 2. The final Pareto front generated by FCE-STRM-MEA III, contrasted with other algorithm versions

The final set of nondominated solutions generated by FCE-STRM-MEA III is better distributed within the feasible region of the Pareto front, in relation to the reference algorithm (RFRC). As RFRC features no encapsulation, this result was anticipated by (18). It is interesting to observe that the suggested algorithm is also superior when compared to PCAM [15], a multiobjective evolutionary identification tool that employs implicit BB exploitation.

## VII. CONCLUSIONS

The paper presents a novel technique for explicit building block exploitation, in the context of NLP model generation. The approach is based on an original extension of the schema theory proposed in [12], that has been specifically tailored to handle regressive models.

The suggested regressive schema theory is used to compute the expected number of schema instances at generation  $t+1$ , whilst using only information available at the current iteration  $t$ . This one step ahead prediction is

valid under both cut point crossover and terminal wise mutation, applied sequentially on the selected parent trees. Computing the transmission probabilities does not require the specific construction of offspring individuals. In addition, applying RST allows the effective use of encapsulation exclusively on trees with slim chances of inherent survival. These last two aspects help reduce the overall computational cost of the algorithm.

By using the suggested RST, the authors have built a sound theoretical model of schemata transmission. This allowed the rigorous demonstration of encapsulation's superiority in propagating a higher number of regressor instances than in the case where protection against genetic operators is not considered. The two layers of encapsulation, consisting in differently altering regressor node selection probabilities in the case of crossover and mutation respectively, is also an original contribution. It allows supplementary insight relative to the way genetic operator interaction influences fit regressor transmission.

The proposed schema based computational instrument may be viewed as a step towards obtaining a general model for genetic programming algorithms, one that would considerably simplify the efforts of the research community towards increasing GP performances.

## ACKNOWLEDGEMENT

The authors would like to thank the Romanian National Center for Programs Management, SICONA research grant - 12100/2008, for their financial support.

- [1] J. Knowles, D. Corne, and K. Deb, MOO Problem Solving from Nature, Springer-Verlag, 2008.
- [2] K. Deb, MOO Optimization using EAs, Wiley&Sons, 2001.
- [3] C.A. Coello Coello, G.B. Lamont, and D.A. Van Veldhuizen, EAs for Solving MOO Problems, 2<sup>nd</sup> ed., Springer-Verlag, 2007.
- [4] J.R.Koza, GP – On the Programming of Computers by Means of Natural Selection, 6<sup>th</sup> ed., MIT Press, 1998.
- [5] O.Nelles, Nonlinear System Identification, Springer-Verlag, 2001.
- [6] J. Madar, J. Abonyi, and F. Szeifert, "GP for System Identification", Proc. of the IEEE Conf. on ISDA, pp 43-48, 2005.
- [7] K. Rodriguez-Vazquez, C.M. Fonseca, and P.J. Fleming, "Identifying the structure of nonlinear dynamic systems using MOO GP", IEEE Trans. On SMC A, 34(4), pp 531-545, 2004.
- [8] E. Vladislavleva, Model Based Problem Solving through Symbolic Regression via Pareto Genetic Programming, <http://arno.uvt.nl/show.cgi?fid=80764>, 2008.
- [9] A. Patelli, and L. Ferariu, Increasing Crossover Operator Efficiency in Multiobjective Nonlinear Systems Identification, Proc of the IEEE Conf on Intelligent Systems, pp 426-431, 2010.
- [10] D.A. Van Veldhuizen, and G.B. Lamont, "Multiobjective Optimization with Messy Genetic Algorithms", Proc of the ACM Symp on Appl Computing, pp.470-476, 2000.
- [11] R.O. Day, Explicit Building Block Multiobjective Evolutionary Computation: Methods and Applications, <http://www.stormingmedia.us/51/5127/A512734.html>, 2005.
- [12] R. Poli, and N.F. McPhee, "General Schema Theory for GP with Subtree Swapping CO II", Evol Comp, 11(2), pp 169-206, 2003.
- [13] N.F. McPhee, and R. Poli, "Using schema theory to explore interactions of multiple operators", GECCO, pp 853-860, 2002.
- [14] A. Patelli, and L. Ferariu, "Regressor Survival Rate Estimation for Enhanced Crossover Configuration", in Adaptive and Natural Computing Algorithms, vol 6593, Springer-Verlag, pp 290-299, 2011.
- [15] L. Ferariu, and A. Patelli, Multiobjective GP for Nonlinear System Identification, in Adaptive and Natural Computing Algorithms, vol 5495, Springer-Verlag, pp 233-242, 2009.

# Vehicle Windscreen Wiper Mathematical Model Development and Optimisation for Model Based Hardware-in-the-Loop Simulation and Control

Jianlin Wei<sup>†\*</sup>, Alexandros Mouzakitis<sup>‡</sup>, Jihong Wang<sup>+</sup>, Hao Sun<sup>+</sup>

<sup>+</sup> School of Engineering, University of Warwick, Coventry CV4 7AL, UK

<sup>‡</sup> Jaguar Land Rover Engineering Centre, Whitley, Coventry, CV3 4LF, UK

\*Corresponding author, Email: J.WEI@warwick.ac.uk, Tel: +442476151580, Fax: +442476 418922

**Abstract**—Hardware-in-the-loop (HIL) simulations have long been used to test electronic control units (ECUs) and software in car manufactures. Accurate Model based HIL simulation (AMHIL) is considered as a most efficient and cost effective way for exploration of new design and development of new products, particularly in calibration and parameterization of vehicle stability controller. The paper presents our recent work in developing an accurate real-time mathematical model for a Jaguar windscreen wiper system, which will be adopted for the HIL tests to the vehicle body control module (BCM). Based on the electro-mechanical engineering principles, the mathematical model of the windscreen wiper system is developed and the unknown model parameters are identified using Genetic Algorithms. Model has been tested by operating at different working conditions, e.g. varied voltage input levels. Real-time implementation of the mathematical model onto dSPACE Ecoline HIL simulator is explained, and the configurations of the laboratory test rig are reported.

**Keywords**—Hardware-in-the-Loop; Real-time Modelling & Control; Genetic Algorithm;

## I. INTRODUCTION

Hardware-in-the-Loop (HIL) simulation has become a well-established verification technology applied in many ECU (Electronic Control Unit) development projects today (ASAM 2009). It provides an effective platform to the development process of the ECU control algorithms with added complexity of the plant under control. The principle of the HIL simulation is illustrated in Figure 1. Rather than using real components/environment to test the control algorithms developed in the ECU, e.g. the real engine as illustrated in the left side of the Figure 1, HIL simulation uses a plant simulator to emulate the real component/environment for testing the ECU. The electrical interface of the ECU is retained, but the real component/environment is substituted intentionally by real-time models. Sensors and actuators are either replaced by full simulated version or they can even attach as original physical load components into the test loop. Accurate real-time simulation of the real component/environment is the vital factor of HIL simulations.

HIL simulators allow developers to validate new hardware and software for automotive solutions, respecting quality requirements and time-to-market restrictions (see Figure 2). Accurate Model based HIL simulation (AMHIL) is considered as a most efficient and cost effective way for exploration of new design and development of new products, particularly in calibration

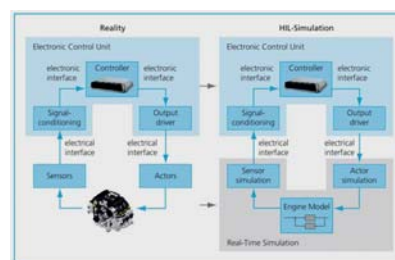


Figure 1. Principle of Hardware-in-the-Loop simulation (ASAM 2009)

and parameterization of vehicle stability controller (dSPACE GmbH 2006). The use of AMHIL simulation can enhance the quality of testing, shorten development time schedule, reduce the cost of reliability test in comparing with using real vehicle components, increase the scope of testing to the level beyond the limitations set by the real vehicle ECU parameters even to test failure conditions, enhance the control functions of the vehicle ECU, provide the possibility of early processing human factors in the loop and support a virtual and real environment for embedded software validation.

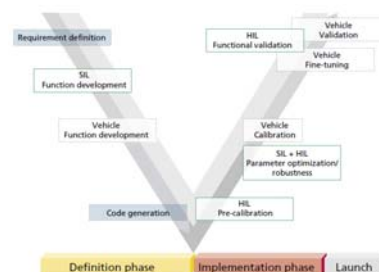


Figure 2. Optimum mix of simulation and test drives to accelerate production process (dSPACE GmbH 2006)

The spirit of AMHIL simulation and test is to use fidelity component and subsystem models to replace their real counterparts and implementation in HIL simulator. To develop high quality vehicle component and subsystem model is a very cost and time consuming task. It requires good knowledge of physics, engineering, mathematics, computer programming and modelling techniques. In this paper, it presents our recent work in developing an accurate real-time mathematical model of the Jaguar windscreen wiper system, which will be dedicated for the HIL tests to the vehicle body control module (BCM).

## II. MODELLING A WIND SCREEN WIPER SYSTEM

The Jaguar windscreen wiper system studied in this paper is the Valeo VM4 Front Wiper System. It has two

major components, which are the DC motor and the Mechanical Linkages (see Figure 3). A Brushed DC motor is used in the wiper system, which is a common type of motors adopted in automotive systems mainly due to its low initial cost. The mechanical linkage system converts the rotational movement of the central shaft, which is driven by the DC motor, to the oscillatory-type motion of the wiper blades.

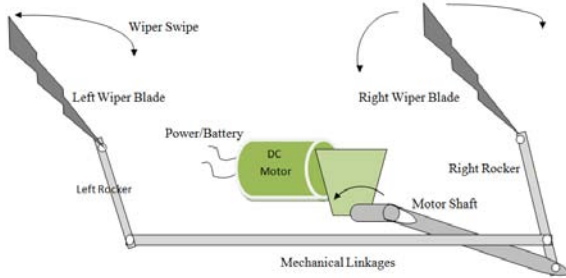


Figure 3. Jaguar's windscreen wiper system

The work described in the paper aims to develop an accurate mathematical model of the wiper system, which can be used to predict the wiper's angle, velocity and acceleration during varied voltage input of the wiper system (see Figure 4).

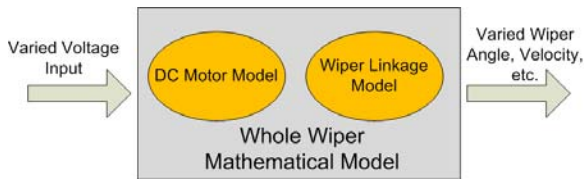


Figure 4. Structure of the wiper system's mathematical model

#### A. Wiper DC motor

The DC motor equipped in the wiper system is a VALEO's continuous motor (X351), which is shown in Figure 5. This type of motor is good for this application as it has the following special features:

- Robust armature thrust system with two ball bearing concept, axial endplay adjustment and a third supporting bearing.
- Robust gear train with rolled shaft and centreline distance adjustment with ex-centre bush concept.

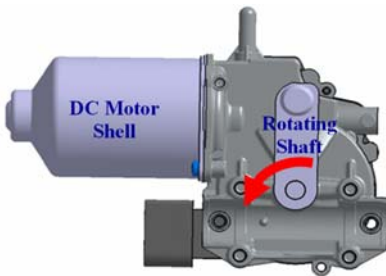


Figure 5. Valeo's Continuous DC Motor used in the wiper system (Valeo Group 2009).

Figure 6 shows the equivalent circuit of the DC motor (Babuska, *et al.* 1999, Aung, *et al.* 2007), which contains electric circuit of the armature in the electrical aspect and also the rotary rotor in the mechanical aspect.

Based on the Kirchhoff's Voltage Law (KVL), the DC motor equation in the electrical aspect can be modelled by the following equation:

$$V(t) - R_a I_a(t) - L \frac{dI_a(t)}{dt} - K_v \frac{d\theta(t)}{dt} = 0 \quad (1)$$

where

$V(t)$  is the voltage source across armature coil at time  $t$ .

$I_a(t)$  is the current across the armature coil at time  $t$ .

$R_a$  is the resistance of the armature.

$L_a$  is the inductance of the armature.

$K_v$  is the velocity constant, which is determined by flux density of permanent magnets, reluctance of iron core of armature, and also number of turns of armature winding.

$\theta(t)$  is the rotational angle of the DC motor rotor at time  $t$ .

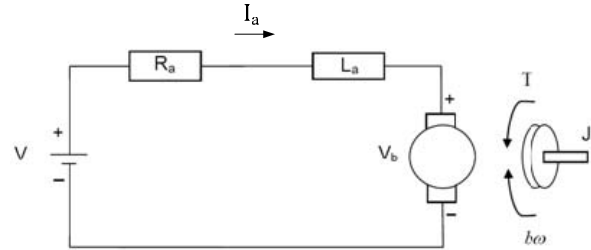


Figure 6. Equivalent circuit of the DC Motor (Babuska, *et al.* 1999)

Based on the torque balance applied to the rotor, the DC motor equation in the mechanical aspect can be modelled by the following equation:

$$K_t I_a(t) - J \frac{d^2\theta(t)}{dt^2} - b \frac{d\theta(t)}{dt} - T_L(t) = 0 \quad (2)$$

where

$K_t$  is the torque constant. Like the velocity constant  $K_v$ , it is determined by flux density of permanent magnets, reluctance of iron core of armature, and number of turns of armature winding;

$J$  is the moment of inertia of rotor;

$b$  is the damping coefficient associated with mechanical rotation of the system.

$T_L(t)$  is the torque due to mechanical load at time  $t$ .

#### B. Wiper Mechanical Linkage

In general, mechanical linkages may be defined as a series of rigid links connected with joints to form one or more closed chains. In our windscreen wiper system, mechanical linkages are used to obtain desired wiper angular velocity and angular acceleration while the motor rotates continuously in one direction. Mechanical linkages are carefully designed and are a very smart means to obtain oscillatory-type motion for wiper blades, thereby enable the direction change of the wiper blades without requiring a change in the direction of DC motor rotation. This proves extremely beneficial because having



bidirectional motion of a motor increases operational complexity and further increases nonlinearities such as friction, backlash and compliance.

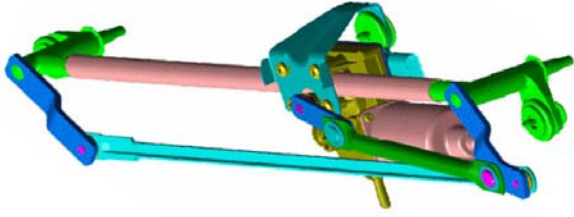


Figure 7. 3D structural diagram of the wiper linkage system (Valeo Group 2009).

3D illustration of the wiper system's linkages can be seen in Figure 7. The linkage have three pivots, which are notated as A, B and C in the Figure 8. The central mount pivot crank (notated as L1) is for the motor crank, wherein the continuous rotating motion is output from the Rotating Shaft. The left and right pivot crank (notated as L6 and L3) are for swinging of the wind screen wipers. The linkage bars (notated as L2 and L5) are for transmitting rotating motion from central mount pivot crank (L1) to the right and left wiper swing cranks (L3 and L6). The dimensions of the linkage system are measured and shown in Table I.

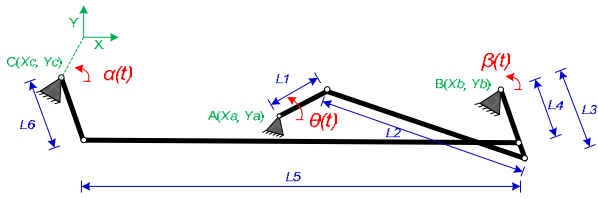


Figure 8. Geometry of the wiper linkage

TABLE I. GEOMETRIC MEASUREMENTS TO THE LINKAGE

$L_1 = 4.50 \text{ cm}$	$X_a = 20.30 \text{ cm}$
$L_2 = 27.00 \text{ cm}$	$Y_a = -2.70 \text{ cm}$
$L_3 = 6.85 \text{ cm}$	$X_b = 47.5 \text{ cm}$
$L_4 = 5.35 \text{ cm}$	$Y_b = 0 \text{ cm}$
$L_5 = 46.70 \text{ cm}$	$X_c = 0 \text{ cm}$
$L_6 = 5.34 \text{ cm}$	$Y_c = 0 \text{ cm}$

From the geometric of the linkage system as shown in Figure 8 and Table I, the angular relationships between the motor driving crank (notated as  $\theta(t)$ ) and the left/right wiper cranks (notated as  $\alpha(t)$  and  $\beta(t)$  respectively) can be presented by the following equations:

$$(X_b + L_3 * \cos(\beta(t)) - X_a - L_1 * \cos(\theta(t)))^2 + (Y_b + L_3 * \sin(\beta(t)) - Y_a - L_1 * \sin(\theta(t)))^2 - L_2^2 = 0 \quad (3)$$

$$(X_c + L_5 * \cos(\alpha(t)) - X_b - L_4 * \cos(\beta(t)))^2 + (Y_c + L_5 * \sin(\alpha(t)) - Y_b - L_4 * \sin(\beta(t)))^2 - L_6^2 = 0 \quad (4)$$

where

$\theta(t)$  : Rotational angle of the DC motor direct driven crank.

$\alpha(t)$  : Rotational angle of the left wiper crank.

$\beta(t)$  : Rotational angle of the right wiper crank.

The animation of the linkage movement between the motor driven crank and the left/right wiper crank is shown in Figures 9. With the smart design of the linkage system, the motion of continuous rotary from the DC motor driving shaft has been transferred to be the oscillatory motion for wiper cranks.

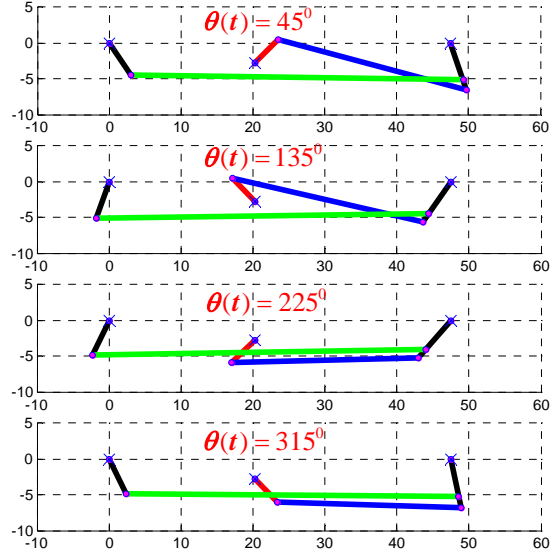


Figure 9. Animation of the linkage movement

### C. Overall wiper model

Assembly the model equations developed in the sections of 2.1 and 2.2, we can have the overall model equations for the whole wiper system, which are shown in Equations:

$$V(t) - R_a I_a(t) - L \frac{dI_a(t)}{dt} - K_v \frac{d\theta(t)}{dt} = 0 \quad (5)$$

$$K_t I_a(t) - J \frac{d^2\theta(t)}{dt^2} - b \frac{d\theta(t)}{dt} - T_L(t) = 0 \quad (6)$$

$$(X_b + L_3 * \cos(\beta(t)) - X_a - L_1 * \cos(\theta(t)))^2 + (Y_b + L_3 * \sin(\beta(t)) - Y_a - L_1 * \sin(\theta(t)))^2 - L_2^2 = 0 \quad (7)$$

$$(X_c + L_5 * \cos(\alpha(t)) - X_b - L_4 * \cos(\beta(t)))^2 + (Y_c + L_5 * \sin(\alpha(t)) - Y_b - L_4 * \sin(\beta(t)))^2 - L_6^2 = 0 \quad (8)$$

$$\omega_\theta(t) = \frac{d\theta(t)}{dt} \quad (9)$$

$$\omega_\alpha(t) = \frac{d\alpha(t)}{dt} \quad (10)$$

$$\omega_\beta(t) = \frac{d\beta(t)}{dt} \quad (11)$$

$$A_\theta(t) = \frac{d\omega_\theta(t)}{dt} \quad (12)$$

$$A_\alpha(t) = \frac{d\omega_\alpha(t)}{dt} \quad (13)$$



$$A_\beta(t) = \frac{d\omega_\beta(t)}{dt} \quad (14)$$

where

$V(t)$ : Voltage source across the wiper system.  
 $I_a(t)$ : Armature current pass through the DC motor.  
 $L_a$ : Armature inductance of the DC motor.  
 $R_a$ : Armature resistance of the DC motor.  
 $K_v$ : Velocity constant of the DC motor.  
 $K_t$ : Torque constant of the DC motor.  
 $J$ : Moment of inertia of the DC motor's rotor.  
 $b$ : Damping coefficient of mechanical system.  
 $\theta(t)$ : Rotational angle of the DC motor direct driven crank.  
 $\omega_\theta(t)$ : Angular velocity of the DC motor direct driven crank.  
 $\alpha(t)$ : Rotational angle of the left wiper crank.  
 $\omega_\alpha(t)$ : Angular velocity of the left wiper crank.  
 $\beta(t)$ : Rotational angle of the right wiper crank.  
 $\omega_\beta(t)$ : Angular velocity of the right wiper crank.  
 $T_L(t)$ : is the torque due to mechanical load.  
 $L_1 \sim L_6, X_a, Y_a, X_b, Y_b, X_c, Y_c$ : Geometric measurements of the linkage system as presented in Table I.

### III. MODEL PARAMETER IDENTIFICATION VIA GENETIC ALGORITHM (GA)

In the wiper model developed in the section 2.3, the dimensional parameters of the linkage system (e.g. length of the linkage bar  $L_1 \sim L_6$ ) are measured. However, the parameters of the DC motor (e.g. moment of inertia of the DC motor's rotor) are not known. In this section, the unknown model parameters are identified. Genetic Algorithms (GAs) have been chosen as a tool for the model parameters' identification.

GA is a robust search mechanism based on the principle of population genetics, natural selection and evolution (Goldberg 1989). It has been proved that GA is a robust optimization method for the parameter identification problems (Karr, *et al.* 1993, Rashtchi, *et al.* 2006, Jia, *et al.* 2008). As a numerical method, the solutions obtained by GA are not mathematically oriented and the GA possesses an inherent robustness according to the design problem specifications. GA is a stochastic search method that mimics the metaphor of natural biological evolution. Basically, the algorithms operate on a population composed of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selection individuals according to their level of fitness in the problem domain and breeding them together (for details of Genetic Algorithms, see (Goldberg 1989). This process leads to the evolutions of the populations, which are better suited to their environment than the individuals that they were created from, just as in natural adaptation. Genetic algorithms can be divided into two categories, the binary GA and the real-value GA, where the GA works with different representations of the variables. Corresponding to the number of the subpopulations in each generation, the GA can also be divided into the single-population GA (SGA) and the multi-population GA

(MPGA). A real-value signal population GA is adopted in identification of the wiper model coefficients, the property setting of which is shown in Table II.

TABLE II. PROPERTY SETTING OF REAL-VALUE SGA

Name of the GA properties	Value of the GA properties
Number of generations	100
Number of individuals per subpopulations	20
Generation gap	0.9
Fitness Assignment	Rank-base fitness assignment
Selection	Stochastic universal sampling
Recombination	Discrete recombination
Crossover rate	0.7
Mutation	Real-value mutation
Mutation rate	0.0625
Reinsertion	Fitness-base reinsertion (global reinsertion)
Reinsertion rate	1.0

As an optimisation method, GA has the advantage that it is able to reach an optimal solution without benefit of explicit knowledge about the problem area. The only existing criterion to evaluate the quality of an individual is the fitness value of this individual. Hence, the evaluation function should provide the required information to GA. Evaluation functions of many forms can be used in a GA, subject to the minimal requirement that the function can map the population into a partially ordered set. As stated, the evaluation function is independent to the GA but greatly dependent on the problem itself. Traditionally, the statistical expectation of Mean Squared Errors (MSE) is widely used to construct fitness functions, which is reasonably effective for static optimisation cases. However, this fitness function may not give a convergent solution for the cases of the dynamic processes. After a number of different fitness functions are studied, it has been found that one particular fitness function gives the algorithm a good convergence and leads to more accurate solutions, which is the sum of absolute values of errors (Wei, *et al.* 2004, 2007, and 2011). The fitness function for identifying the unknown parameters of the wiper model is represented in following equations.

$$fitness = \frac{1}{N} \sum_{t=0}^N |T_{Pulse}(t) - \hat{T}_{Pulse}(t)| \quad (15)$$

where

$T_{Pulse}(t)$ : measured park switch pulse period at time stamp t

$\hat{T}_{Pulse}(t)$ : model simulated park switch pulse period at time stamp t

$N$ : number of data samples

In the wiper system, a park switch sensor is equipped to measure the time when the wiper finish a cycle. A measured park switch signal against the angular values of

the left and right wiper cranks is shown in Figure 10, wherein the wipers are rotated in steady state with 13.5 Volt input.

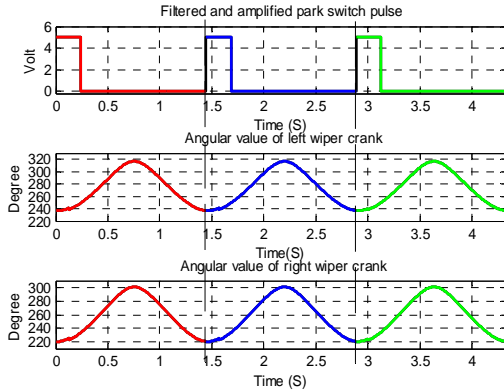


Figure 10. Measured Park Switch against the angular values of left and right wiper cranks.

The park switch pulse period is the time difference between the time stamp of each rising edge of the pulse. An illustration of how the park switch pulse period is calculated is given in Figure 11.

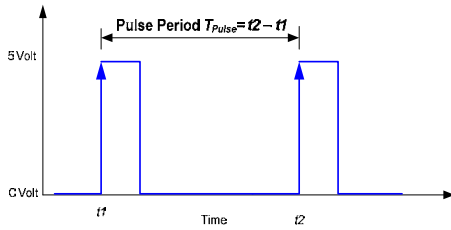


Figure 11. An illustration of the period's calculation of the park switch signal

Following the schematic of model parameters identification as shown in Figure 12, the model parameters are identified in Table III, wherein the back-emf constant  $K_v$  and torque constant  $K_t$  are treated to be the same value as it assumes no electromagnetic losses in the motor. Load torque varies slightly with angle variations. However, it is considered as a constant at this stage of modelling study to simplify the identification process. This will be changed to a time function by using the value identified as the base value in our next stage of research.

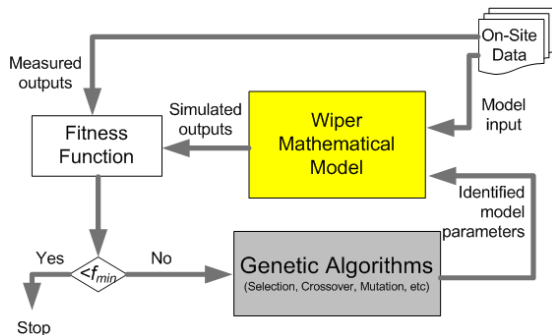


Figure 12. Schematic of the model parameters' identification

TABLE III. IDENTIFIED MODEL PARAMETERS VIA GA

$R_a = 0.9540 \text{ Ohm}$	$J = 0.3938 \text{ N.m.S}^2/\text{Rad}$
$L_a = 0.0152 \text{ H}$	$b = 6.1447\text{e-}005 \text{ N.m.S/Rad}$
$K_v = 0.2746 \text{ Volt.S/Radian}$	$T_L = 0.5279 \text{ N.m}$
$K_t = 0.2746 \text{ N.m/Amp}$	

#### IV. MODEL REAL TIME IMPLEMENTATION IN HiL FACILITY

The model developed in Section 2 is implemented in a real time platform for HiL tests. In nowadays market, there are many real time HiL platforms, e.g. PI Shurlok and NI-PXI. The dSPACE Ecoline simulator is selected as the real time platform due to its powerful calculation ability but yet easy configurations. The simulator is mainly composed of a DS1006 process board (2.6GHz) and DS2211 HiL I/O interface board. Experimental set-up of the wiper test rig is shown in Figure 13.

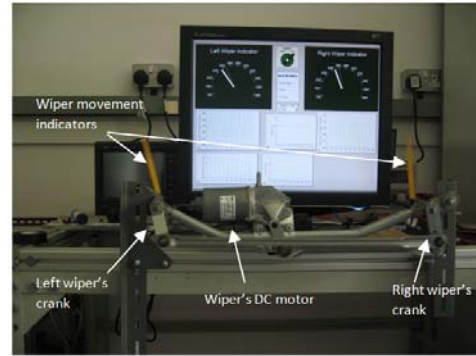


Figure 13. Windscreen Wiper test rig

The procedure for implementing the wiper model into the dSPACE real-time platform is schematized in Figure 14. The wiper system's mathematical model developed in Section 2 is firstly implemented in Simulink (Matlab) using the graphical dataflow programming language. Via the Real-time Workshop from Simulink, the model is further compiled to be real-time code (C language) using the target language compiler of rti1006.tlc. The real-time code of the model is then downloaded into the dSPACE platform to run in real-time.

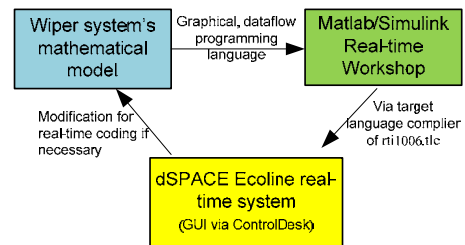


Figure 14. Schematic of model implementation in dSPACE real-time platform

Based on ControlDesk, Graphical user interfaces (GUIs) are developed for communications with the dSPACE platform, wherein controls and measurements to the test rigs take place in real-time. A screen-shots of the GUIs developed is illustrated in Figure 15, which

presents ramp down and up changes of the voltage input to the wiper from 15 Volt to 10 Volt.

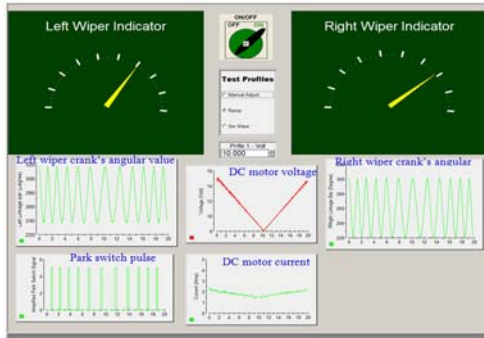


Figure 15. Ramp down and up changes of the voltage input to the wiper

Example experimental results of the wiper model executed in real-time are shown in Figures 16 and 17, wherein the voltage inputs of the wiper are 13.5 Volt.

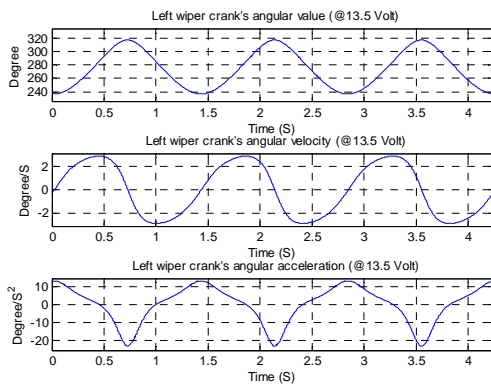


Figure 16. Model simulated outputs in real-time at 13.5 Volt – left wiper crank

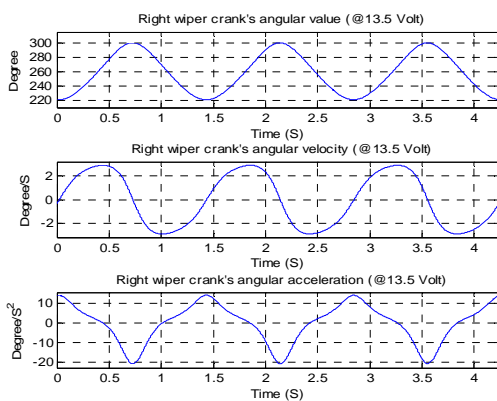


Figure 17. Model Simulated outputs in real-time at 13.5 Volt – right wiper crank

## V. CONCLUSION AND FUTURE WORK

This paper presents the overall process of developing a mathematical model for AMHIL tests. The modelling procedure is based on electrical and mechanical characters of the systems and engineering principles. The unknown model parameters are then identified using Genetic Algorithms with the criteria of time alignment of the model simulated outputs and the measured outputs. The

fitness function is constructed aiming at minimising the error between model simulated park switch pulse and the measured park switch pulse. With the identified model parameters, the mathematical model of the wiper system is then implemented in real-time for the HIL tests. The dSPACE Ecoline simulator is adopted as the real-time platform for the implementation. The experimental results together with test facilities and environment are given in the paper. The project is still on-going within the research group. The wiper system has not been placed the blades and the glass screen, which is in the process of preparation and going to be ready for the whole system model refine and simulation. The paper is to report the work progress and demonstrate the procedure. Also, the new velocity and position sensors are recently installed, which will give extra measured information for our further work. The model will be eventually integrated into the vehicle test system to replace the real components for more cost effective vehicle reliability and robustness test.

## ACKNOWLEDGMENT

The authors would like to thank the support from the AWM/ERDF Birmingham Science City Energy Efficiency and Demand Reduction project.

## REFERENCES

- [1]. ASAM, "Programmers Guide Version 1.0.0", *ASAM AE HIL Application Programming Interface for ECU Testing via Hardware-in-the-Loop Simulation*, 2009
- [2]. Aung, W. P., "Analysis on Modelling and Simulink of DC Motor and Its Driving System Used For Wheeled Mobile Robot", *World Academy of Science, Engineering and Technology* 32, 2007
- [3]. Babuska R., and Stramigioli S., "Matlab and Simulink for Modelling and Control", <http://www.dsc.tudelft.nl>, Control Laboratory, Delft University of Technology, 1999.
- [4]. dSPACE GmbH, "Strategic Use of HIL and SIL", *dSPACE NEWS*, PP16, Volume 3, 2006.
- [5]. Goldberg, D.E., *Genetic Algorithms in Search, Optimisation, and Machine Learning*, Addison-Wesley Publishing Company, Inc, USA, 1989.
- [6]. Karr, C. L., "Genetic Algorithms for Modelling, Design, and Process Control," in *Proc. 2nd Int. Conf. Inf. Knowl. Manage.*, Washington, DC, 1993, pp. 233–238.
- [7]. Jia, N., J. L. Wei, J. Misztal, J. Wang, K. Nuttall, H. Xu, and M. L. Wyszynski, "HCCI Engine Modeling for Control with Parameter Identification using Gas", *11th Mechatronics Forum Biennial International Conference*, Limerick, Ireland, June 2008.
- [8]. Rashtchi, V., E. Rahimpour, and E. M. Rezapour, "Using a genetic algorithm for parameter identification of transformer R-L-C-M model," *Electr. Eng.*, vol. 88, no. 5, pp. 417–422, 2006.
- [9]. Valeo Group, "VM4-Family Front Wiper Motor Technical Presentation", [www.valeo.com](http://www.valeo.com), June, 2009
- [10]. Wei, J. L., J. Wang, and Q. H. Wu, "Study of Fitness Function in Identifying Unknown Parameters of Dynamic Processes using Genetic Algorithms", *Proc. of Postgraduate Research Conference in Electronics, Photonics, Communications and Networks, and Computing Science*, pp. 123-124, Hertfordshire UK, 2004.
- [11]. Wei, J. L., J. Wang, and Q. H. Wu, "Development of a Multi-Segment Coal Mill Model Using an Evolutionary Computation Technique", *IEEE transaction on Energy Conversion*, Vol. 22, Digital Ref. 0885-8969, 2007.
- [12]. Wei, J.L., Jihong Wang, and Q. H. Wu, "Aggregated Power System Load Area Models – Comparison of Three Different Approaches", *International Journal of Modelling, Identification and Control*, PP363-PP377, Vol. 12, No. 4, 2011.

# *Fault Classification of Reciprocating Compressor Based on Neural Networks and Support Vector Machines*

M. Ahmed, S. Abdusslam, M. Baqqar, F. Gu, A.D. Ball  
University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK  
Corresponding author: M.Ahmed@hud.ac.uk

**Abstract:** Reciprocating compressors play a major part in many industrial systems and faults occurring in them can degrade performance, consume additional energy, cause severe damage to the machine and possibly even system shut-down. Traditional vibration monitoring techniques have found it difficult to determine a set of effective diagnostic features due to the high complexity of the vibration signals because of the many different impact sources and wide range of practical operating conditions.

This paper focuses on the development of an advanced signal classifier for a reciprocating compressor using vibration signals. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been applied, trained and tested for feature extraction and fault classification.

The accuracy of both techniques is compared to determine the optimum fault classifier. The results show that the model behaves well, and classification rate accuracy is up to 100% for both binary classes (a single fault present in the compressor) and multi-classes (three faults present).

**Keywords:** *Fault Diagnosis, Reciprocating Compressor, Artificial Neural Networks, Support Vector Machine.*

## I. INTRODUCTION

The use of reciprocating compressors in industry has been widely reported, as has the urgent need for effective condition monitoring, which can accurately detect and diagnose the condition of the compressor see, for example [1].

The vibration signal from a reciprocating compressor contains non-linear characteristics (e.g. due to the impacts resulting from the movement of the suction and discharge valves), and features extracted from the time, frequency and envelope domains of these signals can be used to reliably assess the health of the system. Unfortunately, not all the extracted features are equally useful in trouble-shooting, and experience has shown that even the most useful features are seldom used in the most effective way. In particular the interactions between and among features are not fully considered or even ignored [1] which may undermine the accuracy of diagnosis when the features employed are synergetic.

In this paper Support Vector Machines (SVMs) have been applied to a real compressor with single and multiple faults. It has been claimed that SVMs have four important advantages over the more traditional

ANN. First and most important, is that SVM training uses the powerful mathematical technique of global optimized solutions and so has largely eliminated a major irritant of ANNs: convergence to local maxima and minima [2]. Second the simple geometric interpretation available for SVMs has proved very useful in extending its application to new areas and theoretically can give a sparse solution – that is the solution for the lowest number of entries [3]. Third, during training, the SVM uses structural risk minimization which permits the software designer to allow for sparseness of data and which can lead to a better performance for SVMs than ANNs [4]. Fourthly, it has become clear that SVM is relatively very efficient when dealing with large classification problems (very large feature spaces), because the process of linearization means that the number of dimensions is less important with SVMs than with conventional classifiers [5]. This has the important benefit that the number of features that can be considered for fault diagnosis may be larger than could be used for ANNs.

However, it has also been pointed out that SVMs have a number of less satisfactory features: limited speed both in training and testing, extensive memory requirements, the solutions while geometrically simple can be algebraically complex, and the design of SVMs is not yet anywhere near optimal [6].

The SVM is a binary classifier it compares only two things at a time [7]. This means that if there are  $N$  items to be compared there will  $N*(N-1)/2$  comparisons. Thus, in a real situation there will usually be a huge number of comparisons to be made. This is made worse by the parallel necessity to miss nothing of consequence when taking measurements and to ensure all possible useful features are recorded. But not all features are equally informative about the condition of the machine, and to increase the speed and accuracy of the classifier feature selection and extraction should be limited to those features useful for classification [4-5].

Comparative studies of SVMs and ANNs in fault detection with simple two-class problems (healthy or defective) found that the SVM out-performed the ANN alone in classification accuracy, while performance of the SVM and performance of the ANN combined with a Genetic Algorithm were not significantly

different. However, it was claimed the training time for the SVM was substantially less than required by the ANN, and that the SVM was 100% successful [8].

## II. VIBRATION DATA AND FEATURES

### A. Datasets

Vibration datasets were collected from accelerometers attached near the inlet and outlet valves on the first and second stage cylinder heads of a two-stage, single-acting Broom Wade TS9 reciprocating compressor. The test rig is shown in Figure 1. The compressor delivers compressed air at between 0.55 MPa and 0.8 MPa to a horizontal air receiver tank with a maximum working pressure of about 1.38 MPa. The driving motor was a three phase, squirrel cage, air cooled, type KX-C184, 2.5 kW induction motor. It was mounted on the top of the receiver tank and transfers its power to the compressor through a pulley belt system. The transmission ratio was 3.2:1, so the crank shaft speed was 440 rpm when the motor ran at its rated speed of 1420 rpm. The air in the first cylinder was compressed, passed to the higher pressure cylinder via an air cooled intercooler. When the air pressure in the storage tank reached a prescribed value, a diaphragm pressure switch switched off the electrical current to the motor. The cylinder pressures, temperatures and rotational speed were measured simultaneously with the vibration for comparison. The measured data was then fed, via a

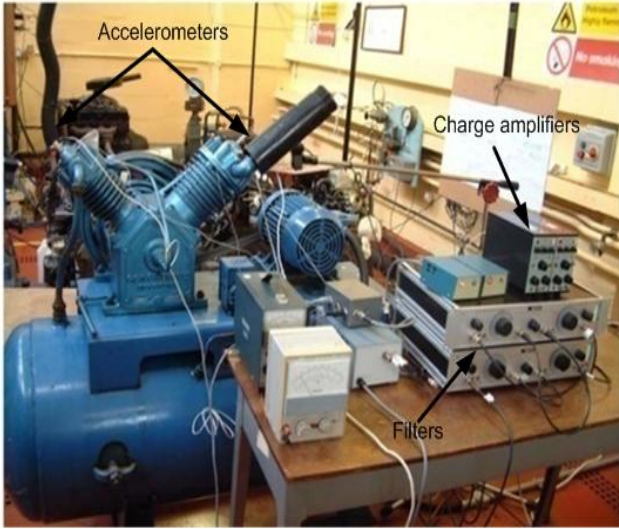


Figure 1 Test rig system

data acquisition system to a computer for further signal conditioning and storage.

Three common faults (loose drive belt, a leaky valve in the high pressure cylinder and a leak in the intercooler) were seeded separately into the reciprocating compressor. The performance of the compressor was monitored with only one fault present at a time. Four sets of experiments were conducted one for normal operation and one for defective operation with each fault. The signal from each channel consisted of 30642

samples at a frequency of 62.5 kHz, total sampling time 0.49 seconds which is more than three working cycles of the compressor. Each data set was divided into 12 segments (bins) of 1024 samples.

### B. Detection Features

The aim was to use signal processing to extract statistical features from the time, frequency and time-frequency domains which are useful for the detection and diagnosis of the seeded faults.

### C. Waveform Features from Time Domain

The features extracted from the vibration signal obtained from the accelerometer on the high pressure cylinder were: root mean square (RMS), peak factor, variance, skewness, kurtosis, range, histogram lower bound (HLB), histogram upper bound (HUB) and entropy. The first five of these are well known so only the last three are defined here:

$$\text{Lower bound} = \min(x) - \frac{1}{2} \frac{\max(x) - \min(x)}{N-1} \quad (1)$$

$$\text{Upper bound} = \max(x) + \frac{1}{2} \frac{\max(x) - \min(x)}{N-1} \quad (2)$$

$$\text{Entropy} = -\sum_{i=1}^N p_i \log p_i \quad (3)$$

Where  $p_i = \frac{x(i)}{\sum_{j=1}^N x(j)}$  and  $\sum_{i=1}^N p_i = 1$ , since  $N$  is the number of samples.

### D. Waveform Features from Frequency Domain

The Fast Fourier Transform (FFT) was used to transform the time-domain signal into the frequency domain from which the spectral features were obtained. The vibration spectra in Figure 2, show a number of discrete components mainly from the compressor working frequency, 7.6Hz, and its harmonics, up to 120 orders. The amplitudes vary slightly but significantly between the different faults, but it was difficult to find a simple set of features to separate the cases completely. Thus the amplitudes of these components were taken as a candidate feature, and different harmonics were used for each trial run. Thus, the resultant was a matrix of spectral features, with  $n$  harmonics and  $s$  the number of samples.

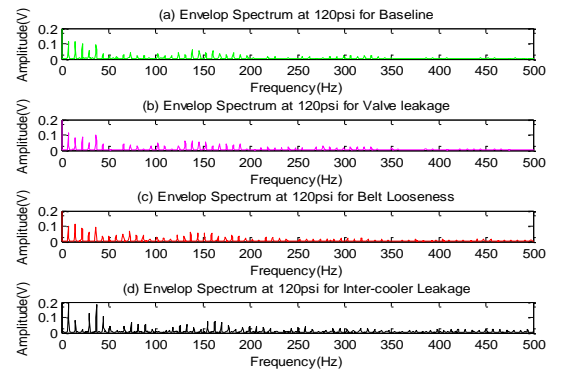


Figure 2 Spectra of compressor vibration for healthy case and three seeded faults



### III. Probabilistic Neural Network

The PNN is a type of supervised neural network introduced by Specht in 1989 and used mainly for classification based on of Bayes optimal decision rule [9]:

$$h_i c_i f_i(x) > h_j c_j f_j(x) \quad (4)$$

where  $f_i(x)$  and  $f_j(x)$  are the probability density functions for data classes  $i$  and  $j$ ;  $h_i$  and  $h_j$  are the prior probabilities;  $c_i$  and  $c_j$  are misclassification data classes. Thus a vector  $x$  is classified into class  $i$  if the product of all the three terms is greater for data class  $i$  than for any other data class  $j$  not equal to  $i$ . In most applications, the prior probabilities and costs of misclassifications are treated as being equal as far as the density functions are concerned. In implementing neural network architecture, a PNN consists of an input layer, a pattern layer, a summation layer and a competitive output layer. This architecture is illustrated in Figure 3.

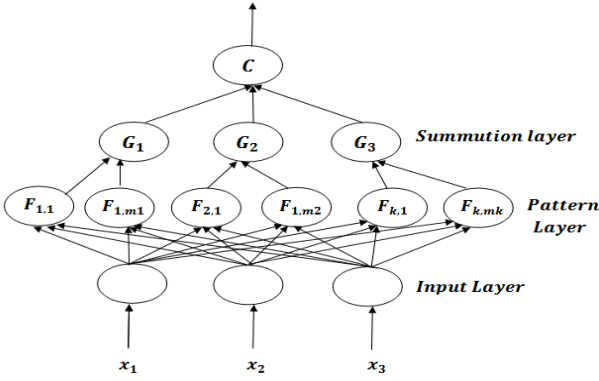


Figure 3. Architecture of a PNN classifier

In recent years, PNN has been widely used in different fields such as pattern recognition and signal processing and has been recognized as a useful technique for high dimensional classification problems. In addition it also is used in CM for differentiating different faults and degrees of fault severity [10].

The PNN is considered much faster than other algorithms such as a Multi-Layer Perceptron (MLP) neural network used in [11] during the training process, which is simply to select a kernel function and its smoothing parameter when solving a linear equation set.

#### A. Pattern Layer

For each training cycle there is one pattern node. For classification the pattern node produces a product of the input pattern vector  $x$  with a weight vector  $w_i$  such that  $Z_i = x \cdot w_i$ , (where both  $x$  and  $w_i$  are normalized) and performs a non-linear operation on  $Z_i$  before outputting its activation level to the summation node. The non-linear operation is  $\exp[(Z_i - 1)/\sigma^2]$ .

#### B. Summation Layer

The summation layer receives the outputs from the pattern layer related to a given class. It sums the inputs from the pattern layer that matched that class from which the training pattern was selected.

$$\sum i e^{[-(w_i - x)^T(w_i - x)/2\sigma^2]} \quad (5)$$

#### C. Output Layer

The output nodes have two input neurons. These units produce binary outputs, associated with two different categories ( $\Omega_s, \Omega_r$   $s \neq r$ ,  $s, r = 1, 2, \dots, q$ ) using the classification principle:

$$\sum i e^{[-(w_i - x)^T(w_i - x)/2\sigma^2]} > \sum j e^{[-(w_j - x)^T(w_j - x)/2\sigma^2]} \quad (6)$$

The outputs have only a single weight  $c$ , given by the loss parameters, the prior probabilities and the number of training patterns in each category. Accordingly, the weight is the ratio of *a priori* probabilities, divided by the ratio of samples, and multiplied by the ratio of losses. These were developed using non-parametric techniques for estimating multivariate or univariate probability density functions from random samples. The  $i^{\text{th}}$  pattern neuron in the  $k^{\text{th}}$  group computes its output using a Gaussian Kernel of the form:

$$F_{k,j}(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{(-\frac{\|x - x_{k,j}\|^2}{2\sigma^2})} \quad (7)$$

Where  $x_{k,i}$  is the centre of the kernel, and  $\sigma$  is a spread parameter which determines the size of the kernel. The summation layer of the network computes the approximation of the conditional class probability function through a combination of the previously computed densities as follows:

$$G_k(x) = \sum_{i=1}^{m_k} \omega_{ki} f_{ki}(x), \quad k \in \{1, \dots, k\} \quad (8)$$

Where  $m_k$  is the number of pattern neurons of class  $k$ , and  $\omega_{ki}$  are positive coefficients satisfying,  $\sum_{i=1}^{m_k} \omega_{ki} = 1$ , pattern vector  $x$  belongs to the class that corresponds to the summation unit with maximum output.

### IV. SUPPORT VECTOR MACHINES

In describing the SVM emphasis is on the engineering and physics. If required, details of the mathematical methods can be found in, e.g.[5, 12-13].

Consider Figure 4, showing only two kinds of training samples:  $\bullet$  and  $\blacksquare$ . Where  $\bullet$  represents healthy and  $\blacksquare$  represents faulty.  $H$  is the classifier hyperplane dividing the two groups of samples;  $x_1$  and  $x_2$ , are the data points closest to  $H$ ;  $H_1$  and  $H_2$  are parallel to  $H$  and pass through  $x_1$  and  $x_2$  respectively. Consider a planar classification task where, optimally, the set of vectors should be separated by the hyperplane without error. The distance separating the closest points of the two classes (distance between  $H_1$  and  $H_2$ ) is defined as the margin [14]. The task is to maximize the margin



(minimise the error bound) to give best performance. Note that this problem is linear.

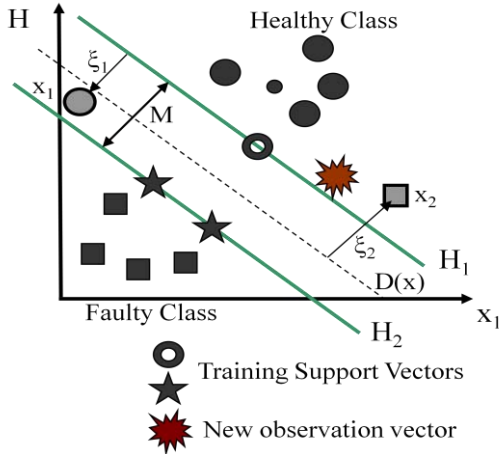


Figure 4 Classification of binary classes using SVM

In standard form the separating hyperplane must satisfy the following constraints:

$$y_i (w \cdot x_i + b) \geq 1 \quad i = 1, 2, \dots, n \quad (9)$$

Where:  $x_i$  is the set of training samples,  $w \cdot x_i$  is the dot product,  $n$  is the number of samples,  $b$  is a scalar measure of the distance of  $H_2$  from the origin, and  $w$  is the normal vector to the hyperplane. Here the samples are assumed be in only one of two classes: healthy or faulty. For the healthy class  $y_i = +1$ , and faulty class,  $y_i = -1$ .

However, in most real situations such an ideal hyperplane does not exist. To find the optimum solution the standard technique is to relax the constraints on (9) by introducing a slack variable,  $\xi_i$  ( $\geq 0$ ). This slack variable is said to represent the noise in the system. The solution to this problem requires the application of advanced but relatively well-known mathematical techniques. The calculation is converted into the equivalent Lagrangian dual problem and the learning task is reduced to minimizing the primal Lagrangian with respect to  $w$  and  $b$ :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i (w \cdot x_i + b) \quad (10)$$

Where  $\alpha_i$  are Lagrangian multipliers.

Finding the optimal values for  $\alpha_i$  allows  $w$  to be expressed in terms of  $\alpha_i$  which allows the solution of (10) to be found. The optimal values for  $\alpha_i$  give the decision function:

$$f(x) = \text{sgn}(\sum \alpha_i y_i (x_i \cdot x_j + b)) \quad (11)$$

This paper refers to a linear problem in which the training samples,  $\bullet$  and  $\blacksquare$ , were separable both in the original input space and in the feature space (hyperspace). However, with multiple dimensions, the

features in the original input space will not normally be separable. Nevertheless a suitable choice of a so-called kernel function to be used in the decision function will separate the features in hyperspace.

$$f(x) = \text{sgn}(\sum \alpha_i y_i (\varphi(x_i) \cdot \varphi(x_j) + b)) \quad (12)$$

The importance of this is that the analysis performed in hyperspace becomes linear. The kernel function is written  $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ . There are now standard kernel functions and this paper uses the very popular polynomial function [15]:

$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^p. \quad (13)$$

## V. IMPLEMENTATION

In this work, the experiments were performed using data from the reciprocating compressor test rig, described above, and computer implementation was conducted in MATLAB.

Figure 5 shows a block flow diagram of a multi-class SVM based fault diagnosis system which consists of three sections: data acquisition, feature extraction and selection, and training and testing for fault diagnosis.

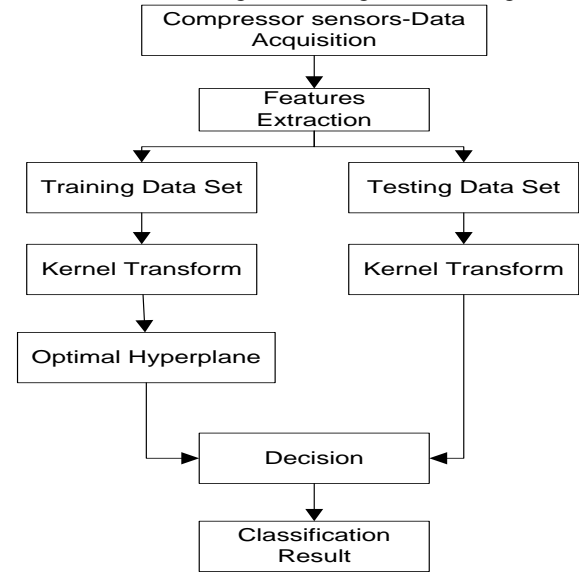


Figure 5 Flow chart of SVM based monitoring

Baseline features were extracted to form a healthy vector feature and faulty conditions created as a vector. A target vector was created the same length as the data vectors. Both data vectors and target vector were divided into two subsets of equal size by taking every other vector value, of which one was for training the SVM and the other for testing. In this particular work a feature selection technique ranks the extracted features and the most important are used as input features. Finally, the SVMs are trained and used to classify the machinery faults.

For comparison, four sets of SVMs have been studied to evaluate the effectiveness of different types of features to calculate the classification rate. The first

two are for the time-domain feature based SVM, the other two is for the frequency-domain feature based SVM.

## VI. RESULTS AND DISCUSSION

Table 1 presents classification results obtained for the SVMs using features extracted from the frequency-domain. There were a total of 120 peaks in the frequency spectrum and each one was a possible feature. In each table there is a column headed “number of features”, the 15 or 20 or other number of features are those which gave the best result. The table includes performance of SVM classifier with a binary class using features from the frequency-domain, and performance of the SVM classifier with multiple classes using features from the frequency-domain.

Number of input features from the frequency domain	Classification success rate % binary class,	Classification success rate % multiple class
15	92.36	83.33
20	85.42	72.92
30	93.75	72.92
45	93.75	82.64
50	94.44	84.03
60	88.33	73.61
75	89.56	79.86
85	84.72	74.31
100	85.45	74.31
120	86.80	71.53

Table 1 Performance of SVM classifier: features from the frequency-domain, single and multiple classes

Table 2 presents results obtained for previously in exactly corresponding situations using a PNN. A comparison shows the PNN is more successful when smaller numbers of features are used, but less successful with larger numbers of features. Interestingly, overall the PNN was more successful than the SVM both at detecting the presence of a single fault (leaky valve) 98.61% compared to 94.44%, and detecting the presence of the three faults, 95.83% compared to 84.03% .

Number of input features from the frequency domain	Classification success rate % binary class,	Classification success rate % multiple class
10	84.72	81.94
15	84.72	81.94
20	91.67	87.70
30	95.83	93.75
45	95.83	93.75
50	97.92	95.14
60	98.61	95.14
65	98.61	95.83
75	88.89	84.03
80	81.25	77.78
85	79.17	72.92
100	71.53	61.81
120	68.75	51.39

Table 2 Performance of PNN classifier: features from the frequency-domain, binary and multiple classes

Table 3 present classification results for binary class fault detection obtained with the SVMs using features extracted from the time-domain. As explained and listed above, nine features were extracted and these were used in different combinations to detect the presence of a single fault (binary classifier) or three faults (multiple classifier). To avoid the need for an extra column in the tables it is stated here that the number of ways of selecting  $n$  features ( $1 \leq n \leq 9$ ) from nine is  ${}^9C_n$ , e.g. there are 126 ways of selecting five features from nine, 126 possible combinations of five features. For example, in the second row of Table 3, features are selected two at a time from the total of nine possible features, there are 36 possible ways of doing this. Of the 36 possible combinations only two (Peak factor and Kurtosis, and Peak factor and Skewness) give the highest classification rate (75%). It can be seen that the SVM was 100% successful in detecting the presence of a single fault when 4, 5, 6 and 7 features were used, but was only 100% successful in detecting the presence of three faults when 5 and 6 features were used.

Number of features used in classification	Number of combinations of features giving highest classification rate	Highest classification success rate %
1	1	50.00
2	2	75.00
3	3	95.83
4	3	100
5	19	100
6	16	100
7	6	100
8	2	100
9	1	91.67

Table 3 Performance of SVM classifier; binary class fault detection using time-domain features

Number of features used in classification	Number of combinations of features giving highest classification rate	Highest classification success rate %
1	1	45.83
2	1	89.56
3	2	93.75
4	3	97.92
5	7	100
6	1	100
7	1	97.92
8	3	95.83
9	1	91.67

Table 4 Performance of SVM classifier; multiple class fault detection using time-domain features

Tables 5 and 6 show the corresponding information for the PNN classifier.

Number of features used in classification	Number of combinations of features giving highest classification rate	Highest classification success rate %
2	7	100
3	15	100
4	35	100
5	35	100
6	21	100
7	7	100
8	1	100
9	1	100

Table 5 Performance of PNN classifier; binary class fault detection using time-domain features

Number of features used in classification	Number of combinations of features giving highest classification rate	Highest classification success rate %
1	1	65.28
2	1	80.56
3	1	93.06
4	3	91.67
5	2	91.67
6	1	91.67
7	3	88.89
8	1	88.89
9	1	83.33

Table 6 Performance of PNN classifier; multiple class fault detection using time-domain features

The PNN classifier is generally more successful than the SVM when only one fault is present. However, the situation is reversed when diagnosing multiple faults when the SVM performed consistently better than the PNN.

## VII. CONCLUSIONS

The PNN clearly performed better than the SVM when diagnosing both the single fault and the three (multiple) faults using features extracted from the frequency-domain.

The performance of the SVM improved considerably when using features extracted from the time-domain. It did not outperform the PNN in the diagnosis of a single fault (binary class) but did much better than the PNN in the diagnosis of three faults, achieving 100% when either five or six features were used.

It should be noted that use of features extracted from the time-domain rather than frequency-domain consistently gave a higher success rate.

## REFERENCES

- [1] B.-S. Yang, *et al.*, "Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines," *Mechanical Systems and Signal Processing*, vol. 19, pp. 371-390, 2005.
- [2] M. Rychetsky, "Algorithms and architecture for machine learning based on regularised neural

networks and support vector approaches," *Sheker-Verlag*, 2001.

- [3] V. Vapnik, "The nature of statistical learning theory," *Springer, New York*, 1999.
- [4] V. Ghate, Dudel, S, "Induction machine fault detection using support vector machine based classifier," *WSEAS Transactions on Systems*, vol. 8, pp. 591-603, 2009.
- [5] A. Widodo, Yang, B-S. , "Support vector machine in condition monitoring and fault diagnosis," *Science Direct, Mechanical Systems and Signal Processing*, vol. 21, 2007.
- [6] J. e. a. Suykens, "Least squares support vector machines," *World Scientific Publishing Company, London*, 2003.
- [7] C. Wei, Chih, H., Lin, J. , "A comparison of methods for multi-class support vector machines," *IEEE Trans. On Neural Networks*, vol. 13, pp. 415-425, 2002.
- [8] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 18, pp. 625-644, 2004.
- [9] S. P. Specht DF, "On fully automatic feature measurement for banded chromosome classification," *Cytometry*, vol. 10, 1989.
- [10] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, pp. 109-118, 1990.
- [11] H. a. Liao, "A comparative study of feature selection methods for probabilistic neural networks in cancer classification," *Proc. 15th IEEE Internat. Conf. on Tools with Artificial Intelligence*, vol. ICTAI, 2003.
- [12] Z. Chen, Lian, X "Fault diagnosis for valves of compressors based on support vector machine," *IEEE Chinese Control and Decision Conference*, pp. 1235 – 1238, 2010.
- [13] S. Gunn, "Support vector machines for classification and regression," *Technical Report Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, Southampton University*, 1998.
- [14] H. P. Chapelle O., Vapnik V.N "Support vector machine for histogram-based image classification," *IEEE Trans on Neural Networks*, vol. 10, pp. 1055-1064, 1999.
- [15] Z. Chen, Lian, X, "Fault diagnosis for valves of compressors based on support vector machine," *IEEE Chinese Control and Decision Conference* pp. 1235 – 1238, 2010.

# Reinforcement Learning based Radio Resource Scheduling in LTE-Advanced

Ioan S. Comşa, Mehmet Aydin, Sijing Zhang,  
Institute for Research in Applicable Computing,  
University of Bedfordshire,  
Park Square, Luton, LU1 3JU, United Kingdom  
E-mails: {Ioan.Comsa, Mehmet.Aydin,  
Sijing.Zhang}@beds.ac.uk

Pierre Kuonen, Jean-Frédéric Wagen,  
Institute of Information and Communication  
Technologies,  
University of Applied Sciences of Western Switzerland,  
Bd. de Pérolles 80, Fribourg, CH-1705, Switzerland  
E-mails: {Pierre.Kuonen, Jean-Frederic.Wagen}@hefr.ch

**Abstract**—In this paper, a novel radio resource scheduling policy for Long Term Evolution Advanced (LTE-A) radio access technology in downlink acceptance is proposed. The scheduling process works with dispatching rules which are various with different behaviors. In the literature, the scheduling disciplines are applied for the entire transmission sessions and the scheduler performance strongly depends on the exploited discipline. Our method provides a straightforward schedule within transmission time interval (TTI) frame. Hence, a mixture of disciplines can be used for each TTI instead of the single one adopted across the whole transmission. The grand objective is to bring real improvements in terms of system throughput, system capacity and spectral efficiency (operator benefit) assuring in the same time the best user fairness and Quality of Services (QoS) capabilities (user benefit). In order to meet this objective, each rule must to be called on the best matching conditions. The policy adoption and refinement are the best way to optimize the use of mixture of rules. The Q-III reinforcement learning algorithm is proposed for the policy adoption in order to transform the scheduling experiences into a permanent nature, facilitating the decision-making on which rules will be used for each TTI. The IQ-III reinforcement learning algorithm using multi-agent environments refines the policy adoption by considering the agents' opinions in order to reduce the policy convergence time.

**Keywords:** *LTE-A, TTI, scheduling rule, policy adoption, Q-III learning, IQ-III reinforcement learning, multi agent systems*

## I. INTRODUCTION

The increase of mobile data usage and the growing demands for new applications (e.g., Massively Multiplayer Online Game, mobile television, web browsing, File Transfer Protocol, video streaming, Voice over Internet Protocol, push-to-talk and push-to-view) have motivated 3<sup>rd</sup> Generation Partnership Project (3GPP) to work with Long Term Evolution (LTE) (3.9 Generation in Mobile Phones (G)) and LTE-A (4G), the latest standards of cellular communication technologies.

Although the previous technologies, such as Global System for Mobile Communications (GSM)/Enhanced Data rates for GSM Evolution (EDGE) (2G/2.5G) and Universal Mobile Telecommunications System/High Speed Downlink/Uplink Packet Access (UMTS/HSxPA)

(3G/3.5G) account at present for over 85% of all mobile subscribers, LTE will provide enhanced performance in comparison with the other mentioned ones.

Evolved Universal Terrestrial Radio Access Network (E-UTRAN), the LTE radio access network, offers important benefits for users and operators [1]: performance and capacity, flexibility, self-configuration and self-optimization, improved cell capacity, reduced latency and mobility. These advantages would not have been possible without some aggressive performance requirements for Physical Layer (PHY) and Medium Access Control (MAC) layers such as [1], [2]: access techniques (Orthogonal Frequency Division Multiple Access (OFDMA)/Single Carrier Frequency Division Multiple Access (SC-FDMA) [4],[5] for downlink/uplink), time and frequency division duplexing, Multiple-In Multiple-Out (MIMO) systems, smart antennas, spectrum efficiency and flexibility and intelligent management of radio resources.

Radio resource management includes transmission power management, mobility management and radio resource scheduling or packet scheduling (PS) [3]. The packet scheduling is a process where the radio resources are assigned to each user in order to offer the requested services in an efficient way by respecting the QoS requirements. Each packet is scheduled in every TTI, a time window used to transmit the user requests and to respond them accordingly.

The packet scheduler is the entity where the resource allocation is made in both time domain (TD) and frequency domain (FD) [6]. The scheduling decision is taken by the following entities: Channel Quality Indicator (CQI) manager, Link Adaptation (LA) unit and Hybrid - Automatic Repeat reQuest (H-ARQ) module [7].

## II. RELATED WORK

Three types of scheduling methods have been proposed in LTE: *dynamic* (MAC and Radio Link Control levels), *persistent* (Packet Data Convergence Protocol and Radio Resource Control levels) and *semi-persistent* [8], [9]. The dynamic scheduler adapts the quantity of resources assigned to users according to their instantaneous channel conditions. While the dynamic

scheduling supports full signaling information, the persistent scheduling is used in order to reduce the signaling overhead for real-time applications [10]. The semi-persistent approach uses both dynamic and persistent scheduling [9]. In [11], it is shown that either of the persistent and semi-persistent scheduling offers a better performance for VoIP traffic when compared with the dynamic scheduling from the cumulative distribution function (CDF) parameter point of view. However, for data oriented applications, it is necessary to have the channel state information, and thus dynamic scheduling provides the best match.

Four main scheduling rules have been proposed in the literature. They are: round-robin (RR), proportional fair (PF), maximum channel/interference (C/I) or maximum rate and delay limited capacity. However, all of these metrics and other derivative ones provide a trade-off between throughput and fairness. For instance, Ramli *et al* [12] presents the performance of packet scheduling algorithms for the downlink single carrier systems. The following metrics for video streaming services are analysed: RR, max-rate, PF, maximum-largest weighted delay first (M-LWDF) and exponential/PF. M-LWDF and EXP/PF algorithms outperform the other ones in terms of the system throughput and RR algorithm provides the best user fairness [12]. RR, PF and max-rate algorithms are compared in [13] from the QoS parameter point of view. It is demonstrated that the PF metric is the best solution for real time applications.

An important issue is how to find the best scheduling metric for the minimum transmit power. The novel minimum transmit power-based packet-scheduling algorithm (MP) is proposed to solve this problem [14]. This method is compared with RR, minimum transmit (MT) and PF metrics for real time (RT) and non-real time (NRT) traffic types in downlink acceptance. The MP algorithm achieves the best average user throughput and the best average cell throughput, provides the best performance in terms of the trade-off between cell throughput and fairness for a minimum transmit power [14].

An opportunistic scheduler is presented in [6]. The opportunistic scheduler is used when the queues are infinitely backlogged. Here, the scheduler identifies channel-aware opportunistic scheduling policies by maximizing the sum throughput. The FD scheduling can be done in one stage or in different stages in order to reduce the system complexity. For the joint optimization model, the users are jointly assigned [6], in comparison with the sub-optimal scheduler, which allocates scheduling blocks (SBs) to users with the highest bit rate in the first stage, and the best MCS for each user is determined in the second stage [6]. The simulated annealing (SA) algorithm is proposed to solve the first model complexity problem. The results show that the SA scheduling method offers comparable performances with the optimization model regarding the average total bit rate but the computational complexity is much lower [6].

The problem of FD packet scheduling has been analysed in [15] where the system includes spatial

division multiplexing (SDM) with MIMO techniques. Here two methods with full and half channel feedbacks (both with SDM-PF) are proposed. The method with partial channel feedback has comparable performance with the full channel feedback method in terms of system throughput, but the signaling overhead is reduced by 50%.

It is almost impossible to optimize the scheduling performance when taking into account only the information from one layer. With the knowledge of the information from the other layers, cross-layer resource allocation schemes are proposed in literature. Adaptive token bank fair queuing (ATBFQ) is an efficient cross-layer scheduling method for packet scheduling and resources allocation, which is focused more on efficiency and QoS mechanisms [16]. The packets are queued in service classes together with per-flow queuing information obtained from the IP layer. ATBFQ provides a very good performance in terms of queuing delay, packet dropping rate comparable with RR and score-based scheduling method (SB). SB is an opportunistic scheduling method which aims to maximize the throughput while maintaining the fairness. ATBFQ, SB and RR show comparable performance on the total sector throughput and the fairness [16].

Two methods proposed in [17] are the genetic algorithm (GA) and the sequential linear approximation algorithm (SLAA), both based on cross-layer resource allocation (CLRA) for the downlink multiuser scheduling for three types of traffic: VoIP, variable bit rate (VBR) video and best effort (BE). The cross-layer technique implies the QoS information which is transferred from the traffic controller to the subcarrier and power allocation and the results are feed backed for the scheduling process. The GA method maximizes the sum weighted capacities of heterogeneous traffic at the PHY layer where the weights are obtained from MAC layer. GA offers a global optimal solution and SLAA generates only local optimal solutions. The GA CLRA outperforms SLAA CLRA in terms of system capacity, total throughput for BE traffic, and average traffic delay for video VBR and VoIP services.

LTE-A introduces important features such as coordination between eNBs (LTE-A base station), cognitive radio, capacity and coverage improvement. A cross-layer resource allocation method for inter-cell interference coordination (ICIC) in LTE-A is analysed in [18]. Inter-eNB coordination for ICIC uses the evolutionary games theory to avoid the interference of resource allocation strategies of cells. The particle swarm optimization (PSO) is used to find the best scheduling scheme for each resource block in the multi-cell case. The results are compared with RR, PF, max C/I and M-LWDF scheduling schemes. The cross layer resource allocation improves the system throughput in a remarkable way and the fairness is guaranteed. The time complexity of this new approach is determined with price of anarchy solution. With this parameter, the cross-layer scheme based on potential games and PSO algorithm finds the optimal solution only after 2 or 3 iterations.

In order to support a wider bandwidth, LTE-A introduces the carrier aggregation (CA) technology. Two or more component carriers belonging to one or more different bands can be aggregated. One of the differences between LTE and LTE-A is that in LTE the user can be scheduled only on one component carrier but, in LTE-A the user is scheduled on multiple component carriers (CC). The carrier load balancing and the packet scheduling for multi-carrier transmission are analyzed in [19]. RR is the rule used for TD scheduling. Therefore, each user must be assigned to CCs. There are two methods of load balancing: round robin and mobile hashing (MH). In FD, the PF algorithm is used with two approaches: independent PS per CC (without the characteristics on the other CCs) and cross CC PS (with the statistics from all CCs). For a large number of users, RR cross CC PS and MH cross CC PS assure the best performance in terms of average cell throughput and coverage throughput. Coverage throughput is taken as the 5% worst user throughput, and therefore a low user throughput indicates poor coverage performance. In FD scheduling, the quantized water-filling packet scheduling scheme with carrier aggregation [20] is also used which offers a significant improvement when is compared with the case without CA from the overall delay point of view.

Each of the above-mentioned scheduling methods is used throughout the entire transmission. Each transmission contains a finite number of TTIs. Therefore, the scheduling performance strongly depends on the exploited scheduling rule and on the adopted performance metrics. And then appears naturally the following question: Can different rules be used instead of one rule adopted across the entire transmission? The answer is “yes” and each rule must be matched with the best conditions. It is necessary to adopt a scheduling policy through which every TTI can work with simple rules.

### III. OUR APPROACH

The length of each TTI is set to be one millisecond. It is necessary to take decisions at every one milisecond in order to determine the rule which will be used. Assume that at the beginning of each TTI,  $U$  users will be allocated with  $M$  resources depending on multiple parameters. If the scheduling algorithm will take decision at every TTI about the rule which will be used, the time complexity parameter will be very high. Then, the policy development becomes mandatory. The policy is an ordered sequence of characters and each character represents a simple scheduling rule.

Let us suppose that we have a pool of  $r$  number of dispatching rules and  $n$  number of TTI intervals. Each rule and each TTI are represented by (1) and (2), respectively, where  $[x]$  is the integer part of  $x$ :

$$R_i \in R = \{R_0, R_1, R_2, \dots, R_{r-l-1}, R_{r-l}, R_{r-l+1}, \dots, R_{r-2}, R_{r-1}\} \quad (1)$$

$$TTI_j \in TTI = \{TTI_0, TTI_1, \dots, TTI_{n-s}, TTI_{n-s+1}, \dots, TTI_{n-1}\} \quad (2)$$

$$\text{where } \begin{cases} i = \overline{0, r-1} \\ \lfloor r/2 \rfloor - 1 < l < \lfloor r/2 \rfloor + 1 \end{cases} \quad \text{and} \quad \begin{cases} j = \overline{0, n-1} \\ \lfloor n/2 \rfloor - 1 < s < \lfloor n/2 \rfloor + 1 \end{cases}$$

Then, the policy  $p_k$  can be defined as:

$$p_k = \{R_{r-l-2}, R_0, R_{r-l+1}, \dots, R_l, R_{r-l+4}, R_{r-3}, \dots, R_{r-l-1}, R_{r-l+2}\} \quad (3)$$

where  $p_k$  has  $n$  elements and  $k$  ( $k \in \mathbb{N}$ ) is the communication session index.

Fig. 1 illustrates the principle of FD scheduling for the old and for the new approach (it is chosen the index  $L$  instead of index  $l$  in order to avoid the confusion between  $l$  and  $l$ ). Here, any rule from the pool can be used for the  $k^{\text{th}}$  transmission session in connection with the lot of parameters and circumstances. However, for the new approach (Fig. 1.b), the policy decides which is the selected scheduling rule for each TTI. For instance, the rule  $R_{r-l-2}$  is applied for  $TTI_0$ ,  $R_0$  for  $TTI_l$  and so on.

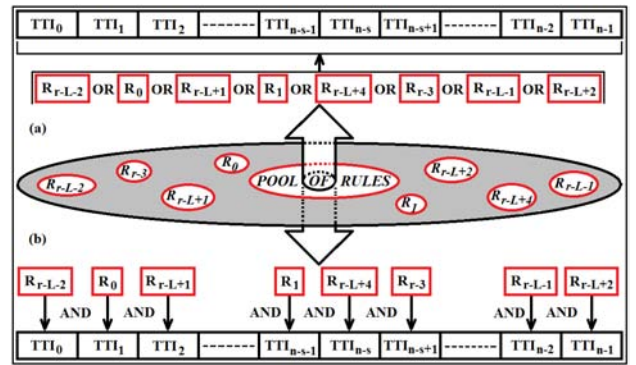


Figure 1. FD scheduling for the old approach (a) and for the new approach (b)

Each rule from the rules set represents an action. We define a state as a group of  $U$  users (packets) and  $C$  conditions (radio sub-channels conditions, QoS bearers). Each TTI represents a matrix with  $S$  columns (symbols) and  $M$  radio resources (sub-carriers). Basically, the TTI matrix is the result of applying the selected rule for a certain state. By performing an action  $R_i \in R$ , the agent will move the simulation from one state ( $TTI_{n-s}$ ) to the next state ( $TTI_{n-s+1}$ ).

The state-rule-TTI performance is evaluated by the utility function  $U(t)$  which can be defined as:

$$U(t) = \text{OperatorBenefit}(t) \times \text{UserBenefit}(t) \quad (4)$$

The idea is to evaluate the utility function for each TTI (local evaluation) rather than for the whole transmission (global evaluation). For each TTI, the utility function is represented in (5).

$$Q(t_j, R_i) = \text{OperatorBenefit}(t_j) \times \text{UserBenefit}(t_j) \quad (5)$$

where  $Q(t_j, R_i)$  is the utility function for  $j^{\text{th}}$  TTI when the  $R_i$  rule is applied. Then, the global utility function is the sum of local utility functions:

$$U(t) = \sum_{j=0}^{n-1} Q(t_j, R_i) \quad (6)$$



The scheduler target is to maximize the utility function. In this case, we maximize this function in two ways. For the global case, the utility function is maximized by using a proper scheduling rule for the entire transmission as is shown in (7). For the local case, the function is maximized using different rules for each state (TTI) in order to maximize the local utilities (8).

$$U'(t) = \max \left( \sum_{j=0}^{n-1} Q(t_j, R_i) \right) \quad (7)$$

$$U''(t) = \sum_{j=0}^{n-1} \max (Q(t_j, R_i)) \quad (8)$$

Without any demonstration, which is not the object of this paper, we have the next inequality:

$$U'(t) < U''(t) \quad (9)$$

With (9), our approach will have a great advantage assuring a greater utility than in the old approach.

The main problems are how to select the proper rules in order to make a local maximizations and how to find the most representative policy such as  $p_k$ . If  $P$  is the policy for a large number of communication sessions, it is represented by:

$$P = \{p_0, p_1, p_2, \dots, p_{k-2}, p_{k-1}, p_k\}, k \in \mathbb{N} \quad (10)$$

The optimization set for the same number of communication sessions is:

$$P^{opt} = \{p_0^{opt}, p_1^{opt}, p_2^{opt}, \dots, p_{k-2}^{opt}, p_{k-1}^{opt}, p_k^{opt}\}, k \in \mathbb{N} \quad (11)$$

Then, for the  $k^{\text{th}}$  transmission session, the optimization function can be defined as:

$$f: P \rightarrow P^{opt} \quad f(p_k) = opt(p_k) = p_k^{opt} \quad (12)$$

We propose two main approaches for the policy adoption and optimization:

1. evolutionary programming with genetic and hyper-heuristic algorithms;
2. dynamic programming with Markovien decision and temporal difference-based learning algorithms.

In this paper, we will analyze the Q-III learning algorithm. In order to achieve this, Fig. 2 illustrates the modality of the policy optimization. The simulation platform will send the information of conditions to the intelligent agent. The perception procedure is able to convert the conditions in the perceptible way for agent. With the learning algorithm, the agent should be able to take decisions, in order to select the proper rule from the rules set corresponding to the received conditions. The results are transferred to the simulation environment through the action procedure where the rule is applied. In the literature, there are a lot of learning algorithms, but a reinforcement algorithm is proposed in [21] which makes the difference from the other methods. The Q-III learning algorithm is able to transform the temporal experiences into permanent nature. So, the experience from the previous transmission sessions can be used for the current transmission session.

The agent receives a reward value from each state. The main issue is to maximize the total reward. This is achievable if the agent will learn the optimal action for each state.

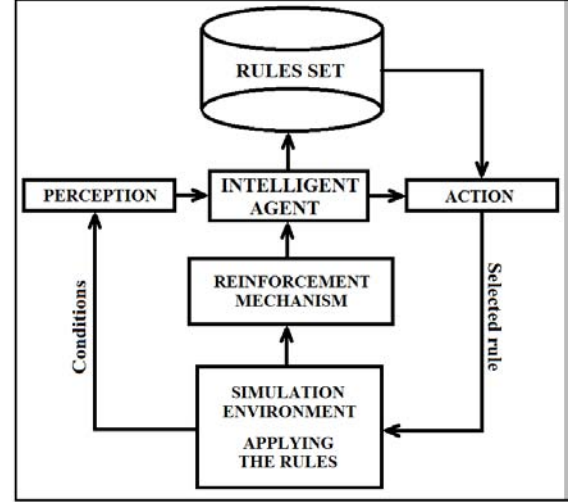


Figure 2. The intelligent agent structure

The Q-III learning algorithm is developed based on the utility function ( $Q$ ) which calculates the quality of state-rule-TTI combination. Definition domains for the  $Q$  function are illustrated in (13) and the  $Q$  values are calculated in (14).

$$Q: \mathbb{R} \rightarrow \mathbb{R} \quad (13)$$

$$Q^n(t, R_i) \leftarrow \alpha_{ii}(Q^o(t, R_i) + \beta(rw + \gamma e(t+1) - Q^o(t, R_i))) \quad (14)$$

where  $i$  is the rule index,  $Q^o(t, R_i)$  and  $Q^n(t, R_i)$  are the old and the new state of  $Q$  values for rule  $R_i$  applied in the  $t^{\text{th}}$  state (TTI),  $\alpha_{ii}$  represents the relationship between  $i^{\text{th}}$  and  $l^{\text{th}}$  rules,  $\beta$  and  $\gamma$  are the learning and discounted rates with  $\beta, \gamma \in (0, 1)$ ,  $rw$  is the reward value of applying the rule  $R_i$  for the  $t^{\text{th}}$  state (TTI), and  $e(t+1)$  is the expected value of state  $t+1$ .

The learning rate expresses how much the oldest information will affect the newest information. For instance, if the learning rate will be 0, the agent will not learn anything. For  $\beta = 1$ , the agent will take into account only the most recent information.

The discount rate determines the importance of future rewards. If the factor is 0 the agent will take into account only the current rewards, if the factor converges to 1, the agent will have a long term high reward.

The Q-III learning algorithm defines the  $e(t+1)$  function with a strong predictor using the past experiences of the agent.  $e(t+1)$  is calculated with Hard c-Means algorithm based on a distance function  $d$  and it is defined by (15).

$$e(t+1) = d(t_j - v_w) = \|t_j - v_w\| = \sqrt{\sum_{i=0}^{r-1} (t_{ji} - v_{wi})^2} \quad (15)$$

where  $\|\cdot\|_{i=0, r-1}$  represents the Euclidean distance between  $t_j^{\text{th}}$  state and  $v_w$  data center where,  $w$  is the dimension of data center and  $i$  is the rule index. The data center is determined using Hard c-Means algorithm expressed in (16).

$$v_{wi} = \left( \sum_{j=0}^u \chi_{wj} t_{ji} \right) / \left( \sum_{j=0}^u \chi_{wj} \right) \quad (16)$$

where  $u$  is the maximum number of TTIs (states), the  $i^{\text{th}}$  rule is applied and  $\chi_{wj}$  is a membership of the  $j^{\text{th}}$  TTI to the  $w^{\text{th}}$  center as is shown in (17).

$$\chi_{wj} = \begin{cases} 1, & \text{for } r = (-1) \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

The  $v_{wi}$  value can be calculated with (18) in order to simplify (16).

$$v_{wi} = T_{wi} / u \quad (18)$$

where  $T_{wi}$  is the cumulative data for the data center with  $i^{\text{th}}$  rule and  $u$  is the number of accumulated data.

For the current transmission time interval, the  $Q$  value can be recalculated and defined as:

$$Q(t, R_i) \leftarrow rw + \gamma \left( \sqrt{\sum_{i=0}^{r-1} (t_{ji} - v_{wi})^2} \right) \quad (19)$$

where  $t_{ji}$  is the current TTI with the  $i^{\text{th}}$  rule.

At the beginning,  $Q(R_i)$  and  $T_{wi}$  are initialized to 0. Then, the first state ( $t$ ) is perceived and analyzed. The rule which has the maximum utility ( $Q$ ) value is selected in order to find the data center closest to the current state. After the rule is executed, a new state is created ( $t+1$ ). The reinforcement mechanism produces the reward value. If the agent takes the right decision regarding the rule that will be applied to the current state, the RM will send a reward value equal to  $-1$ . Otherwise, the reward value will be 1. For  $rw = -1$ , the agent will recalculate the cumulative data  $T_{wi}$  using the relation (20) and then, all the quality values are updated using (14), (15) and (18). If  $rw = 1$ , the data centers are not updated.

$$T_{wi} \leftarrow T_{wi} + t \quad (20)$$

To obtain a smooth optimization of sequences such as  $p_k$ , each algorithm will require the execution process over a large number of scenarios. The searching space will be  $M \times U \times R \times T$ , where  $M$  is the number of radio resources,  $U$  is the number of users,  $R$  the number of rules and  $T$  the number of TTIs. The workload can be estimated as several days of running the programs because a large number of scenarios will be considered for each algorithm. The solution represents the use of Q-III reinforcement learning in multi-agent systems based on parallel and distributed computing to obtain a wider and more comprehensive experimentation.

### Q-III learning algorithm in multi agent systems

It is necessary to extend the Q-III learning algorithm for multi-agent environments in order to solve the convergence capacity problem.

In this paper, a Q-III learning algorithm for multi agent-systems is proposed in order to improve and to hasten the convergence capacity for obtaining a policy such as  $P^{\text{opt}}$  described in (11). The learning method is based on influence value paradigm. The main idea behind

of this proposal is that the behavior of each agent is influenced by the opinions of the other agents. Let us suppose that we have a number of  $N$  agents defined by the set represented below:

$$A_a \in A = \{A_0, A_1, \dots, A_N\} \quad (21)$$

where  $A_a$  is the current agent. If the agent  $A_a$  applies the rule  $R_i$  to  $j^{\text{th}}$  state ( $TTI_j$ ) and all the other agents like this action, they must praise agent  $A_a$ . If the agent  $A_a$  continues to take the right decisions of applying the same rule for each TTI, the others will become accustomed and over time they will stop to praise  $A_a$ . The idea is to praise  $A_a$  inversely proportional with the number of times when the proper rule is applied.

The influence value for the agent  $A_a$  in a group of  $N$  agents is defined by the following expression:

$$I_a \leftarrow \sum_{b=1, b \neq a}^N \theta_a(b) * O_a(b) \quad (22)$$

where,  $\theta_a(b)$  is the influence rate of agent  $A_b$  over agent  $A_a$  when applying a rule, and  $O_a(b)$  is the opinion of agent  $A_b$  regarding the applied rule by the agent  $A_a$ .

The opinion value is calculated based on the reward values obtained by the agent at the moment of evaluation not in the past like in  $Q$  value cases. So, the reward value for the agent  $A_a$  can be defined using the rewards values obtained in the past plus the quality of rule-TTI value and the predicted  $Q$  value for the next TTI using Hard c-Means algorithm:

$$rwv_b \leftarrow rw_b + e_b(y) - Q(t, R_i(b)) \quad (23)$$

where  $R_i(b)$  is the rule index applied by the agent  $A_b$  at the  $t^{\text{th}}$  moment. With (24), the opinion value is represented as follows:

$$O_a(b) = \begin{cases} rwv_b * oc(t, R_i(a)) & \text{if } rwv_b \leq 0 \\ rwv_b * (1 - oc(t, R_i(a))) & \text{if } rwv_b > 0 \end{cases} \quad (24)$$

where  $oc(t, R_i(a))$  is the occurrence index representing the number of times when the agent  $A_a$  applies the same rule  $R_i$  to the transmission session. Thus, in the IQ-III learning algorithm for multi-agent environments, the quality of state-rule-TTI values for the agent  $A_a$  can be calculated by extending the relation (14).

$$Q^n(t, R_i(a)) \leftarrow \alpha_{il} (Q^{n-1}(t, R_i(a)) + \beta(rw_a + \gamma e_a(t+1) - Q^{n-1}(t, R_i(a)) + I_a)) \quad (25)$$

The Algorithm 1 will explain in more detail the proposed algorithm by extending the Q-III learning algorithm.

---

#### Algorithm 1. IQ-III learning algorithm for the agent $A_a$

---

1. **Initialize**  $Q(t, R_i(a))$ ,  $T_{wi}$ ,  $v_{wi}$  to 0;
2. **Initialize** the first state to  $TTI_0$  ( $t = TTI_0$ ) for all agents
3. **Repeat** the following steps until the transmission is over
4.     Perceive the current TTI ( $t$ )
5.     Select the rule  $R_i$  with the maximum  $Q$  value
6.     Apply the rule  $R_i$

```

7.    Consider the next TTI(t+1) and assign the reward  $rw_a$ 
8.    If  $rw_b$  is (-1) then
9.        Recalculate the cumulative data  $T_{wi}$ 
10.        $T_{wi} \leftarrow T_{wi} + t$ 
11.       and update the data center
12.        $v_{wi} = T_{wi} / u$ 
13.   else do not update the data center
14.   Determine the reward  $rw_b$  for the current evaluation
15.    $rwv_b \leftarrow rw_b + e_b(t+1) - Q(t, R_i(b))$ 
16.   Monitor the other agents actions  $(A_0, A_1, \dots, A_N)$ 
17.   For  $A_b =$  all the agents except  $A_a$  do
18.        $O_a(b) = \begin{cases} rwv_b * oc(t, R_i(a)) & \text{if } rwv_b \leq 0 \\ rwv_b * (1 - oc(t, R_i(a))) & \text{if } rwv_b > 0 \end{cases}$ 
19.   end for
20.   Monitor the opinions of all agents
21.    $I_a \leftarrow \sum_{b=1, b \neq a}^N \theta_a(b) * O_a(b)$ 
22.   Update the Q values with the following rule
23.    $Q^n(t, R_i(a)) \leftarrow$ 
        $\alpha_{ii}(Q^{n-1}(t, R_i(a)) + \beta(rw_a + \gamma e_a(t+1) - Q^{n-1}(t, R_i(a)) + I_a)$ 
24.   where  $e_a(t+1) = d_a(t_j - v_w) = \|t_{ji} - v_{wi}\| = \sqrt{\sum_{i=0}^{r-1} (t_{ji} - v_{wi})^2}$ 
25.    $t \leftarrow t+1$ 
26. end

```

#### IV. CONCLUSIONS

In this paper, a novel radio resource scheduling methodology has been proposed. In comparison with other existing methods in the literature, our approach is able to increase the utility function and to minimize in the same time the trade-off between the operator and user benefits. This is achievable if users are scheduled with different scheduling disciplines during each TTI. The policy adoption implements the Q-III reinforcement learning algorithm, where the past scheduling experiences are learned by the agent. The influence value of IQ-III reinforcement learning provides a very good support to increase the convergence capability.

#### REFERENCES

- [1] Ericsson, White Paper, "LTE – an introduction", June, 2009.
- [2] Motorola, White Paper, "Long Term Evolution: A Technical Overview".
- [3] S. Hussain, "Dynamic Radio Resource Management in 3GPP LTE", Blekinge Institute of Technology, January, 2009, pp. 11-45
- [4] R. Kwan, C. Leung, J. Zhang, "Multiuser Scheduling on the Downlink of an LTE Cellular System", in Hindawi Publishing Corporation, 27 May 2008.
- [5] J. Zyren, W. McCoy, "Overview of the 3GPP Long Term Evolution Physical Layer, in Freescale Semiconductor", July, 2007
- [6] M. Aydin, R. Kwan, J. Wu, J. Zhang, "Multiuser Scheduling on the LTE Downlink with Simulated Annealing", in VTC2011 Spring Budapest, May, 2011.
- [7] S. Lu, Y. Cai, L. Zhang, J. Li, P. Skov, C. Wang, Z. He, "Channel-Aware Frequency Domain Packet Scheduling for MBMS in LTE", National Basic Research Program of China.
- [8] E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, S. Parkvall, "Key features of the LTE radio interface", Ericsson Review No. 2, 2008
- [9] D. Vinella, M. Polignano, "Discontinuous reception and transmission (DRX/DTX) strategies in Long Term Evolution(LTE) for voice-over-IP traffic under both full-dynamic and semi-persistent packet scheduling policies", Aalborg University, November 20th, 2009.
- [10] C. Leung, R. Kwan, S. Hamalainen, W. Wang, "LTE/LTE-Advanced Cellular Communication Networks", in Journal of Electronic and Computer Engineering, June 2010, pp. 1-10.
- [11] J. Puttonen, N. Kolehmainen, T. Henttonen, M. Moisio, "Persistent Packet Scheduling Performance for Voice-over-IP in Evolved UTRAN Downlink", in IEEE Xplore, 2009.
- [12] H. Ramli, R. Basukala, K. Sandrasegaran, R. Patachaianand, "Performance of Well Known Packet Scheduling Algorithms in the Downlink 3GPP LTE System", in Proceedings of the 2009 IEEE 9th Malaysia International Conference on Communications, December, 2009.
- [13] H. Ramli, K. Sandrasegaran, R. Basukala, R. Patachaianand, T. Afrin, "Video streaming Performance under wellknown packet scheduling algorithms", in International Journal of Wireless & Mobile Networks (IJWMN) Vol. 3, No. 1, February 2011.
- [14] J. Song, G. Gil, Kim, "Packet-scheduling algorithm by the ratio of transmit power to the transmission bits in 3GPP LTE downlink", in EURASIP Journal on Wireless Communications and Networking, July, 2011.
- [15] S. Lee, S. Choudhury, A. Khoshnevis, S. Xu, S. Lu, "Downlink MIMO with frequency-domain packet scheduling for 3GPP LTE", in the IEEE INFOCOM proceedings, 2009
- [16] F. Bokhari, H. Yanikomeroglu, W. Wong, M. Rahman, "Cross-layer resource scheduling for video traffic in the downlink of OFDMA-based wireless 4G Networks", in EURASIP Journal on Wireless Communications and Networking, 2009.
- [17] Nan Zhou, Xu Zhu and Yi Huang, "Genetic algorithm based cross-layer resource allocation for wireless OFDM networks with heterogeneous traffic", in the 17th European Signal Processing Conference, Glasgow, August, 2009.
- [18] Z. Lu, Y. Yang, X. Wen, Y. Ju, W. Zheng, "A cross-layer resource allocation scheme for ICIC in LTE-Advanced", in Journal of Network and Computer Applications, December, 2010.
- [19] Y. Wang, K. Pedersen, T. Sørensen, P. Mogensen, "Carrier load balancing and packet scheduling for multi-carrier systems", in IEEE Transactions on Wireless Communications, vol.9, no.4, April, 2010.
- [20] Y. Chung, Z. Tsai, "A quantized water-filling packet scheduling scheme for downlink transmissions in LTE-Advanced systems with carrier aggregation", National Taiwan University.
- [21] M. Aydin, E. Öztemel, "Dynamic job-shop scheduling using reinforcement learning agents", in Robotics and Autonomous Systems, vol. 33, December, 1999, pp. 169–178.

# The extraction of characteristic quantity of shallow defects in pulsed magnetic flux leakage signal

<sup>1</sup>Junbiao Fei, <sup>1</sup>Xianzhang Zuo, <sup>2</sup>Yunze He, <sup>2</sup>Guiyun Tian, <sup>1</sup>Tao Zhang

<sup>1</sup>Department of Electric Engineering, Ordnance Engineering College  
Shijiazhuang, China

<sup>2</sup>School of Electrical, Electronic and Computer Engineering, Merz Court, Newcastle University  
Newcastle upon Tyne, NE1 7RU, England, UK

[feijunbiao@yahoo.cn](mailto:feijunbiao@yahoo.cn)

**Abstract**—In order to enhance the testing ability of pulsed magnetic flux leakage (PMFL) for shallow defects, the method of second difference for dealing with testing signals is put forward, which separates the damp components, namely second differential signals, caused by eddy current effect from the signals in pulsed magnetic flux leakage. From second differential signals, a new characteristic quantity of the depth of shallow defects—peak time  $Pt$  is extracted and the feasibility of the method of second difference is verified. The experimental results show that compared with the value of signal amplitude, the resolution of the depth of defects is higher while testing shallow defects if  $Pt$  is taken as the characteristic quantity of shallow defects and at the same time, the problem of the seriously decreasing resolution of the characteristic quantity of defects caused by the lift-off of the testing probe can be solved.

**Keywords**—pulsed magnetic flux leakage testing; eddy current effect; second differential signals; characteristic quantity of shallow defects; resolution

## I. INTRODUCTION

Pulsed magnetic flux leakage (PMFL) is a new nondestructive testing technology developed recently. Because it combines the features of pulsed eddy current testing technology and those of magnetic flux leakage technology, its potential superior reflects on quantitative evaluation of defects in ferromagnetic materials [1-2]. According to reference [3], for plane conductor with infinite depth, the distribution of its eddy current density decreases exponentially with the increasing distance from the surface of the conductor. The distance penetrated by eddy current into the conductor is called penetrating depth. The penetrating depth is defined as the standard penetrating depth when eddy current density decreases to  $1/e$  of the value of its surface, also called as skin depth. Skin depth is relevant to driving frequency, the electrical conductivity and the permeability of the conductor. Its expression is as follows:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \quad (1)$$

In the expression,  $\delta$  is skin depth (m);  $\omega$  is angular frequency (rad/s);  $\mu$  is permeability (H/m);  $\sigma$  is electrical conductivity (S/m).

According to Fourier inversion formula, pulsed stimulating square-wave signal contains abundant frequent

components in pulsed magnetic flux leakage testing so compared with traditional magnetized way of single-rate exchange, the penetrating depth penetrated by stimulating magnetic field in specimen has much been deepened; at the same time, according to equation(1), the distribution of magnetic field tends near the surface of the magnetic circuit because of skin effect and compared with static magnetic field, the testing sensibility for defects on upper and lower surfaces is higher.

Currently, the scholars at home and abroad study more of the characteristic quantity of the defects in the signals in pulsed magnetic flux leakage testing. Reference [4] makes the qualitative analysis for signals in pulsed magnetic flux leakage at the point of defect. The result shows that compared with the signals in magnetic flux leakage, the signals in pulsed magnetic flux leakage contain more defective information in the time domain and frequency domain. References [5-6] use the method of finite element to make simulation the testing for rectangular defects on pipe and analyses qualitatively the relationship between peak voltage of testing signals and defective depth, width and the lift-off of the sensor. The result shows that signal amplitude is more sensitive to the change of the lift-off value. References [7-8] use the situation where the rectangular defects simulate the corrosion of the pipe to study the relationship between peak time of testing signals and defective depth and the position of upper and lower surfaces.

The paper takes the shallow defects as the research object and in order to enhance the testing ability of pulsed magnetic flux leakage for shallow defects and reduce the influence of lift-off effect to signal characteristic resolution, on the basis of analyzed the features of eddy current effect in pulsed magnetic flux leakage testing, the method of second difference is proposed to separate the damping components of pulsed magnetic flux leakage signals caused by eddy current effect and extract the new characteristic quantity of shallow defective depth.

## II. EDDY CURRENT EFFECT IN PULSED MAGNETIC FLUX LEAKAGE TESTING

The principle of the ferromagnetic materials in pulsed magnetic flux leakage testing is shown in figure 1. The sensor contains a U-shaped magnet yoke, an stimulating coil and a testing probe. The pulsed stimulating signal is a square wave with certain frequency and duty cycle which loads on the stimulating coil around magnet yoke and thus it



can generate the pulsed transient magnetic field in the magnetic circuit. When there are defects in the specimen, the field of pulsed magnetic flux leakage at the point of the defect will change so it changes the transient voltage induced by the testing probe. The defective situation can be known through the analysis of the transient voltage [9-10].

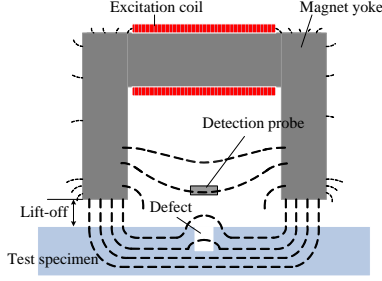
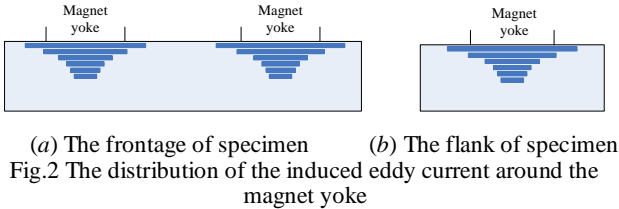
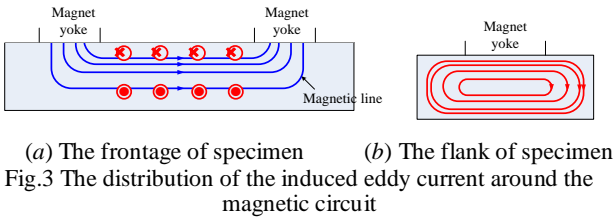


Fig.1 The principle of PMFL testing

Because the simulating magnetic field in pulsed magnetic flux leakage testing is transient magnetic field, the tested specimen will induce eddy current field under the influence of simulating magnetic field when the U-shaped sensor in pulsed magnetic flux leakage tests ferromagnetic materials. The induced eddy current can be seen as the overlap of two eddy current fields. One is shown in the figure 2, centering two magnet yokes respectively and distributing near the upper surface of the specimen; the other is shown in the figure 3, circulating the distribution of the magnetic circuit in the inner part of the specimen [11]. Because the permeability of ferromagnetic materials is more higher, according to the skin effect formula (1) and Maxwell's equations, the penetrating depth of induced eddy current in the specimen is very shallow and at the same time the induced strength is very strong.



(a) The frontage of specimen (b) The flank of specimen  
Fig.2 The distribution of the induced eddy current around the magnet yoke



(a) The frontage of specimen (b) The flank of specimen  
Fig.3 The distribution of the induced eddy current around the magnetic circuit

At the same time, according to reference [12], the depth  $d$  corresponding with the time  $t$  transmitted by pulsed electromagnetic wave in the medium of the conductor can be gained according to equation (2), in which  $\mu_r$  and  $\sigma$  are respectively corresponding permeability and electrical conductivity of the metal conductor.

$$d = \sqrt{\frac{t}{\pi\mu_r\sigma}} \quad (2)$$

It can be known that transmitting speed of pulsed electromagnetic wave in the medium of the conductor is inversely proportional to the product of relative

permeability and electrical conductivity. Taking aluminum and steel as an example, according to the relative permeability of aluminum  $\mu_r=1$ , electrical conductivity  $\sigma=3.4e7$  S/m, the relative permeability of steel  $\mu_r=300$ , electrical conductivity  $\sigma=5e6$  S/m to calculate, transmitting speed of pulsed electromagnetic wave in the medium of aluminum is 44 times than that of steel. In other word, the resolution of the peak time of eddy current density in steel is 44 times than that in aluminum, which indicates peak time of induced eddy current density has very strong resolution on the depth direction in pulsed magnetic flux leakage testing.

### III. THE METHOD OF SECOND DIFFERENCE

According to further analysis, when the method of pulsed magnetic flux leakage is used to test ferromagnetic materials, eddy current effect in the specimen will form damp for the change of magnetic field in magnetic circuit. When there are defects in the specimen, the defects will disturb the flowing pattern of the induced eddy current and thus affect the damping effect of eddy current effect to magnetic field. Because of different characteristics of the disturbance in induced eddy current caused by defects with different characteristics, damping effect of eddy current effect to magnetic field is also different. Therefore, the damping component in magnetic field caused by eddy current effect should contain defective information.

In order to separate the damping component from pulsed magnetic flux leakage signals, the method of second difference is proposed for pulsed magnetic flux leakage signals. Its basic idea is that assuming A and B are two specimens with defects, A is ferromagnetic material and the electrical conductivity of B is 0, whose material attribute, size and the settings of defect is same with A. Using the sensor of pulsed magnetic flux leakage to test A and B, pulsed magnetic flux leakage signals  $B_{za}$  and  $B_{zb}$  are extracted respectively on the same positions above the defects of A and B and then referential signals  $B_{zac}$  and  $B_{zbc}$  are extracted respectively on the position without defects and differential signals  $\Delta B_{za}=B_{za}-B_{zac}$  and  $\Delta B_{zb}=B_{zb}-B_{zbc}$  can be acquired. Apparently, the signal  $\Delta B_{za}$  can be acquired when eddy current effect exists but the signal  $\Delta B_{zb}$  is acquired when there is no eddy current effect. In other word, the damping component contained in  $\Delta B_{za}$  isn't contained in  $\Delta B_{zb}$ . Theoretically,  $\Delta B_{zb}-\Delta B_{za}$  is the damping component separated from pulsed magnetic flux leakage signals, which reflects the variance of eddy current effect caused by the disturbance of defect to damping effect, containing defective information. Here,  $\Delta B_{za}$  is also called first differential signal and  $\Delta B_{zb}$ , regarded as the referential signal of  $\Delta B_{za}$  with second difference, is called second differential referential signal. As for the damping component  $\Delta B_{zb}-\Delta B_{za}$ , it is acquired after pulsed magnetic flux leakage signal  $B_{za}$  conducts twice of difference so  $\Delta B_{zb}-\Delta B_{za}$  is called second differential signal.

### IV. THE CHARACTERISTIC QUANTITY OF SHALLOW DEFECTS IN SECOND DIFFERENTIAL SIGNAL

The software of finite element analysis, ANSYS, is used to calculate the model of pulsed magnetic flux leakage testing shown in figure 4. In order to acquire the first

differential signal  $\Delta B_{za}$  and the second referential signal  $\Delta B_{zb}$ , two groups of the testing specimens are respectively designed with length 150mm, width 100mm, height 10mm. In group one, the relative permeability is  $\mu_r=300$ , electrical conductivity  $\sigma=5e6S/m$ ; in group two, the relative permeability is  $\mu_r=300$ , electrical conductivity  $\sigma=0S/m$ . Each group contains five specimens. Every specimen contains one defect and the length (10mm) and the width (2mm) are the same and the heights are 0.1mm、0.2mm、0.3mm、0.4mm、0.5mm respectively.

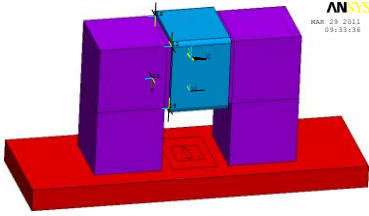


Fig.4 The simulating model of the finite element

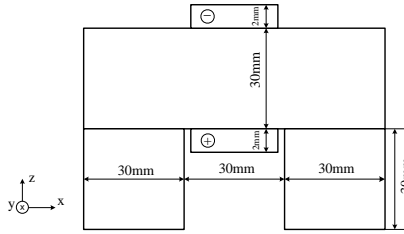


Fig.5 The size of the sensor

The settings of the sensor's size and the direction of coordinates are shown in figure5. The stimulating coil is the copper wire with 400 cycle and the diameter 0.3 mm and the stimulating current with the square wave function increasing exponentially, whose duty cycle is 50%, amplitude 0.3A and frequency 50 Hz, is selected. Because of the features of the time domain in pulsed stimulating signals, the model needs conducting transient analysis and in order to assure the accuracy of the calculated result, multiple load steps also needs setting in the solver. Because transient calculation occupies more resources and longer time, considering the symmetry of waveform of stimulating current, only positive half period of the stimulation is calculated.

Figure 6 shows a typically first differential signal  $\Delta B_{za}$ , second referential signal  $\Delta B_{zb}$  and second differential signal acquired of their variance. The tendency of the signals  $\Delta B_{za}$  and  $\Delta B_{zb}$  is the same, both increasing sharply at first and then the increasing speed slows down. However, the second differential signal increases at first and then decreases. What Figure 7 shows is the signal acquired at the midpoint of the edge of defect's length when the lift-off value is 1mm and the result after dealing with second difference. In the figure, the defects with the depth from 0.1mm to 0.5mm, whose peak time of second differential signal  $Pt$  appears at 279 $\mu s$ , 298 $\mu s$ , 315 $\mu s$ , 331 $\mu s$ , 344 $\mu s$  in sequence, the variance of time corresponding with variance of depth 0.1mm  $\Delta Pt$  is 19 $\mu s$ , 17 $\mu s$ , 16 $\mu s$ , 13 $\mu s$  in sequence. It shows that under the situation where the least depth variance of defects on surface is only 0.1mm, peak time of second differential signal has good resolution when taken as the characteristic quantity of the defect on depth direction,

which corresponds with the features of high resolution of peak time of density of induced eddy on depth direction current in ferromagnetic materials.

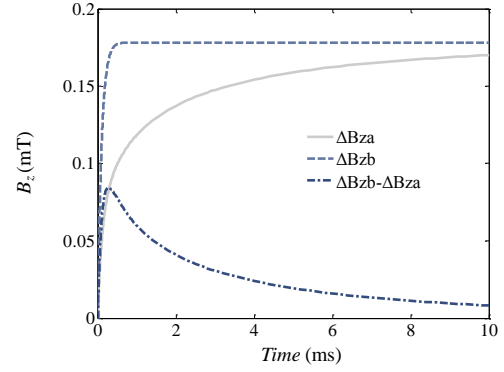


Fig.6 The typical signal in the method of second difference

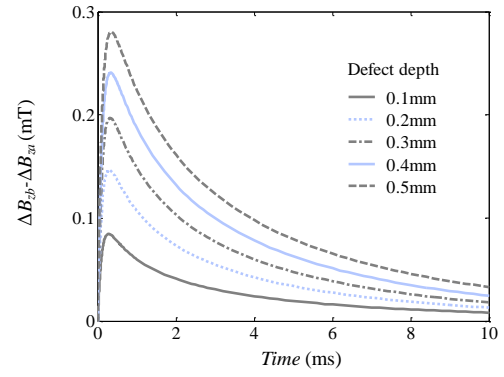


Fig.7 Theoretical second differential signal

However, what worth paying attention to is that it is hard to find the material with electromagnetic attribute in practice, like the specimen B and also hard to use testing method to acquire  $B_{zb}$  and  $B_{zbc}$  and then separate damping component from pulsed magnetic flux leakage signal through  $\Delta B_{zb} - \Delta B_{za}$ . Thus, we consider looking for a signal which can substitute  $\Delta B_{zb}$ . Through the analysis of the pulsed magnetic flux leakage signal  $B_{zb}$  extracted above the specimen B and referential signal  $B_{zbc}$ , we have found that  $B_{zb}$  and  $B_{zbc}$  at different points are all in linear relationship with stimulating signal  $\varphi_e$  and then it indicates that  $\Delta B_{zb}$  is also in linear relationship with  $\varphi_e$ . According to this characteristic, considering making the stimulating signal zoom out in a appropriate proportion through  $\varphi = k \cdot \varphi_e$ , substitute the signal  $\varphi - \Delta B_{za}$  for  $\Delta B_{zb} - \Delta B_{za}$  and then the feasible method of second difference can be gained. Because the existence of induced eddy current impacts damping effect on magnetic field in magnetic circuit overall. After the amplitude of stimulating magnetic field stabilizes, induced eddy current will weaken rapidly and the damping effect will also gradually disappear. Theoretically, if the pulsed width is wide enough, after induced eddy current disappear completely, the waveform of  $\Delta B_{za}$  and  $\varphi_e$  should be in linear relationship and the value of amplitude same with  $\Delta B_{zb}$  but before this,  $\Delta B_{za}$  in general should be less than  $\Delta B_{zb}$ . At the same time, considering from the angle of frequency domain, testing signals under the stimulation of the pulsed square wave should all contain defective information of each frequent component. As for the proportional relationship of signals with the power of different frequency components, it



is mainly decided not by the value of amplitude but by signal's waveform of time domain. In turn, considering from the angle of time domain, we can infer that defective information should mainly be contained in the signal's waveform, and especially the recognition for different defects should be mainly relied on the relationships between different signals' waveforms, not only on the value of amplitude. Therefore, we here mainly consider retaining the information of signal's waveform and take the quotient of peak value of differential signal  $\Delta B_{za}$  in pulsed magnetic flux leakage and that of stimulating signal  $\varphi_e$  as value K and then deduce  $\Delta B_{zb}$ 's approximate value  $\varphi$ .

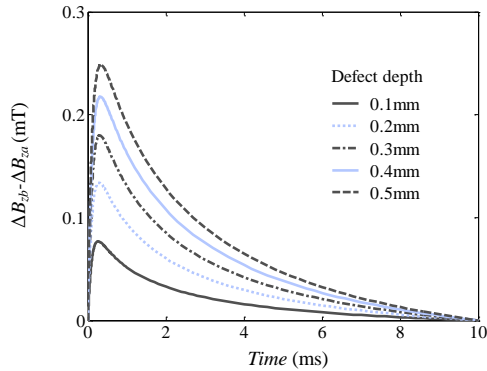


Fig.8 Applied second differential signals

Figure 8 shows the results dealt with pulsed magnetic flux leakage signal by the method of applied second differential signals. The depth of defects varies from 0.1mm to 0.5mm and peak time of their second differential signals appears at 270 $\mu$ s, 290 $\mu$ s, 307 $\mu$ s, 323 $\mu$ s, 336 $\mu$ s in sequence. What variance of time corresponding with variance of depth 0.1mm is 0 $\mu$ s, 17 $\mu$ s, 16 $\mu$ s, 13 $\mu$ s in sequence. Compared with the second differential signals in theory, the resolution of characteristic quantity of defects practically doesn't change; with the defective depth deepening per 0.1mm from 0.1mm to 0.5mm, the increasing proportion of signal's amplitude are 73.1%, 34.7%, 22.3%, 16.3% respectively

and 74.4%, 34.3%, 20.9%, 14.4% respectively in theoretical calculation and the value of amplitude and changing tendency is practically the same with the theory, which indicates the method of second difference is feasible in practical testing.

## V. THE ANALYSIS OF TESTING RESULTS

Build pulsed magnetic flux leakage testing system as figure 9 shows, impact the square wave of pulsed electrical current with the frequency 50Hz, amplitude 0.5A and duty cycle 50% on stimulating coil and test respectively for defects on the artificial surfaces with same length (8mm) and same width (1mm), depth 0.1mm, 0.2mm, 0.3mm, 0.4mm, 0.5mm respectively. Using Hall element testing probe to extract pulsed magnetic flux leakage signals on the midpoint of defective length without lift-off, referential signal is acquired at the point without defects. The mean amplitude of the first differential signal is 938.79mV and peak time of second differential signal is averagely 156.7 $\mu$ s. The variance of mean time is 12.8 $\mu$ s when variance of defective depth is 0.1mm.

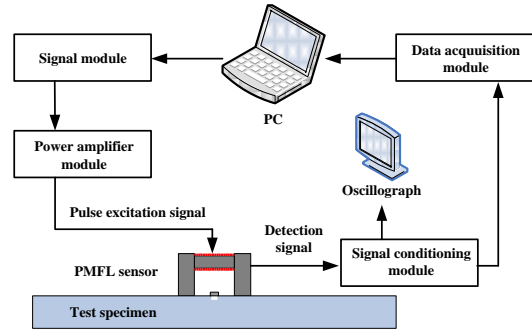


Fig.9 PMFL testing system

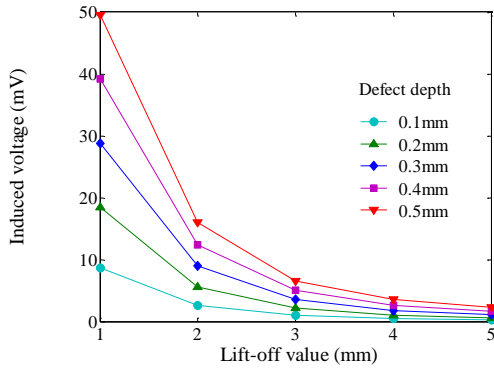
The data results are obtained through first and second difference dealing with the signals with different lift-off value (1mm~5mm), as shown in figure 10, table 1 and table 2.

Tab.1 The resolution of the amplitude of first differential signals

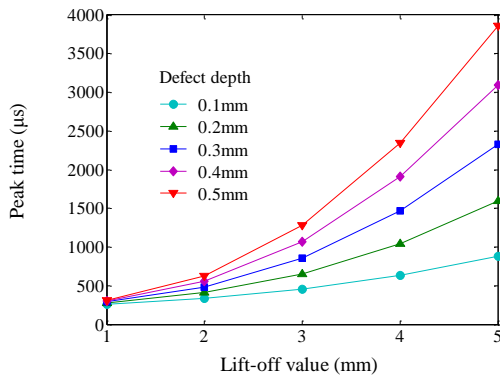
Lift-Off Value(mm)	Variance of Amplitude $\Delta A$ of First Differential Signals in Different Depths (mV)				Variance of Mean Amplitude (mV)	Mean Percentage of Reducing (%)
	0.1mm~0.2mm	0.2mm~0.3mm	0.3mm~0.4mm	0.4mm~0.5mm		
1mm	9.84	10.31	10.41	10.32	10.22	*
2mm	3.06	3.33	3.48	3.56	3.36	67.17
3mm	1.23	1.37	1.47	1.53	1.40	58.26
4mm	0.58	0.69	0.90	0.98	0.79	43.72
5mm	0.34	0.50	0.57	0.62	0.51	35.31

Tab.2 The resolution of peak time of second differential signals

Lift-Off Value(mm)	Variance of Peak Time $\Delta Pt$ of Second Differential Signals in Different Depths ( $\mu$ s)				Variance of Mean Time ( $\mu$ s)	Mean Percentage of Increasing (%)
	0.1mm~0.2mm	0.2mm~0.3mm	0.3mm~0.4mm	0.4mm~0.5mm		
1mm	17	15	12	11	13.8	*
2mm	73	71	72	74	72.5	427.3
3mm	196	204	216	209	206.3	184.5
4mm	409	426	437	440	428.0	107.5
5mm	713	747	764	764	747.0	74.5



(a) Characteristic quantity of first differential signals



(b) Characteristic quantity of second differential signals  
Fig.10 characteristic quantity with different lift-off value

It can be seen in figure 10 that the amplitude  $A$  of first differential signal weakens with the increasing lift-off value and peak time  $Pt$  of second differential signal becomes longer with the increasing lift-off value. Table 1 and Table 2 reflect that the resolution of two characteristic quantities with the increasing lift-off value of testing probe tends to change reversely. The resolution of the first differential signal's amplitude with the lift-off value of the testing probe being 5mm reduces 95% compared with the resolution with the lift-off value of the testing probe being 1mm but the resolution of the peak time of second differential signal increases 3000%; at the same time, the mean amplitude of the first differential signal with lift-off value 1mm is 36.14mV and variance of mean amplitude is 10.22mV. Compared with corresponding value without lift-off value, it decreases 96.15% and 92.07% respectively, which indicates that the amplitude of first differential signal is very sensitive to the lift-off value, especially the testing probe lift-off from 0mm to 1mm, characteristic quantity and resolution decreasing very rapidly.

## VI. CONCLUSION

This paper separates the damping component, namely second differential signal, from the signals in pulsed magnetic flux leakage testing, caused by eddy current effect through the proposed method of second difference and according to the characteristic of eddy current effect in ferromagnetic materials, characteristic quantity of shallow defects is extracted from second differential signal—peak time  $Pt$  of second differential signal. The experimental result shows that amplitude of the differential signal in pulsed magnetic flux leakage is relatively sensitive to the

lift-off of testing probe and its resolution decreases rapidly with the lift-off value increasing but as for new characteristic quantity still has relatively high resolution under the situation where testing probe lifts and the resolution increases with the lift-off values increasing. It provides a new efficient method for pulsed magnetic flux leakage testing for shallow defects.

## REFERENCES

- [1] John W. Wilson, Gui Yun Tian. Pulsed electromagnetic methods for defect detection and characterisation[J]. NDT&E International, 2007, 40, pp. 275-283.
- [2] Sophian A, Tian G Y, Zairi S. Pulsed magnetic flux leakage techniques for crack detection and characterization[J]. Sensors and Actuators A, 2006, 125, pp. 186-191.
- [3] Ali Sophian, Gui Yun Tian, David Taylor, et al. A feature extraction technique based on principal component analysis for pulsed Eddy current NDT[J]. NDT&E International, 2003, 36, pp. 37-41.
- [4] Sophian A, Tian G Y, Zairi S. Pulsed magnetic flux leakage techniques for crack detection and characterization[J]. Sensors and Actuators, 2006, 125, pp. 186-191.
- [5] Tang Ying, Luo Feilu, Pan Mengchun, et al. 3D magnetic field analysis and defect characterization of pulsed magnetic flux leakage field testing[J]. Chinese Journal of Scientific Instrument, 2009, 30(12), pp. 2506-2510. (in chinese)
- [6] TANG Ying, LUO Feilu, PAN Mengchun, et al. Finite element numerical simulation on steel pipe using pulsed magnetic flux leakage testing[J]. Non-destructive testing, 2009, 31(7), pp. 513-516. (in chinese)
- [7] Tang Ying, Pan Mengchun, Luo Feilu, et al. Detection of corrosion in pipeline using pulsed magnetic flux leakage testing[J]. Computer Measurement & Control, 2010, 18(1), pp. 38-39. (in chinese)
- [8] WANG Yunjiang, WANG Xiaofeng, DING Keqin. Width quantification of corrosive defect on pipeline based on pulsed magnetic flux leakage[J]. Journal of Test and Measurement Technology, 2009, 23(5), pp. 390-395. (in chinese)
- [9] Tian Lu Chen, Gui Yun Tian, Ali Sophian, et al. Feature extraction and selection for defect classification of pulsed eddy current NDT[J]. NDT&E International, 2008, 41(6), pp. 467-476.
- [10] Huang Zuoying, Que Peiwen, Chen Liang. 3D FEM analysis in magnetic flux leakage method[J]. NDT&E International, 2006, 39(1), pp. 61-66.
- [11] Javier García-Martín, Jaime Gómez-Gil, Ernesto Vázquez-Sánchez. Review Non-Destructive Techniques Based on Eddy Current Testing [J]. Sensors, 2011, 11(3), pp. 2525-2565.
- [12] J.Blitz, T.S.Peat, The Application of Multi-frequency Eddy Currents to Testing Ferromagnetic Metals[J]. NDT International, 1981, 14, pp. 15-17.

# Detection technology to identify money based on pulsed eddy current technique

Sumin Qian<sup>1</sup>, Xianzhang Zuo<sup>1</sup>, Yunze He<sup>2</sup>, Guiyun Tian<sup>2</sup>, Hong Zhang<sup>2</sup>

<sup>1</sup> Department of Electrical Engineering, Ordnance Engineering College  
Shijiazhuang, China

<sup>2</sup> School of Electrical, Electronic and Computer Engineering, Newcastle University  
Newcastle upon Tyne, NE1 7RU, UK  
qiansumin1988@yahoo.com.cn

**Abstract**—In today's society, the national large denomination currencies are used in paper currency, thus inevitably coming with a high degree of counterfeit banknotes, so for the inspection and recognition of notes, there have been many new methods with technological advances. Electromagnetic detection is a new technology of detection to identify money. This paper introduces the basic principles of pulsed eddy current technology, the inspection equipment design, the experimental results analysis and the characteristic quantity extraction. The results show that, notes detection has a high detection resolution and the potential advantages for development based on the pulsed eddy current technology. It can distinguish between different denominations by further practical design and may count different denominations at the same time.

**Keywords**—pulsed eddy current; notes identification; feature extraction

## I. INTRODUCTION

With the development of society, paper currency has been widely used. Printing technology and anti-counterfeiting technology are constantly increasing, but the technology of manufacturing counterfeit currency is also rising. Therefore, notes testing technology is constantly developing. The current anti-counterfeiting technologies include security line, watermarks, magnetic inks, security signs, the entire department photographic and perspective printing technology. The main identifications consist of fluorescence identification, magnetic ink identification, security line identification, watermark identification, graphic identification, comprehensive identification of paper, infrared identification, identification of light transmission. The application of the various money detectors is generally integrated of several detection techniques<sup>[1]</sup>.

Anti-counterfeiting of security line is in the use of a special symbol in the note to detect whether the bank notes are true or not. Security lines generally include metal lines, opaque plastic line, fluorescence line, microfilm printing lines. For the security line detection, UV irradiation can be used for banknotes. With ultraviolet radiation, a security line will show in different positions of the notes and different security line will appear in different colors. However, under the UV irradiation, security lines of notes with a long time may not be able to show<sup>[2]</sup>. While the metal security thread has magnetism, pulsed eddy current detection methods can be used. There is also a common method which is to test for magnetic ink of notes<sup>[3]</sup>, the ink

of the notes is magnetic ink. In electromagnetic fields, electromagnetic induction signal can be measured by the sensor. However, the magnetic ink is easy to wear, causing the detection of instability.

Most of existing detectors test notes in qualitative detection in which watermarks or security line exist or not. The current counterfeiting currency also has these features, so it is sometimes difficult to distinguish clearly. How the security features of banknotes can be detected in the quantitative resolution has been a hot research. This article focuses on detecting the security line of money with pulsed eddy current technology in the quantitative detection to identify the authenticity of banknotes, and lays the foundation of counting the mixed banknotes.

## II. PULSED EDDY CURRENT TESTING TECHNOLOGY

Eddy current testing is one of the most widely used methods in the evaluation of non-destructive testing for the conductor surface. The principle of eddy current testing method is as follows: when a coil with alternating current closes to the sample, the electromagnetic induction occurs in the alternating magnetic field of the coil and sample, induces the eddy current in conductor and magnetic field is induced by eddy current. When the surface of sample exists defects, it will lead the strength and distribution of eddy current and the induced magnetic field changes the single-frequency alternating current was used in conventional eddy current testing technology and impedance (or voltage) of amplitude and phase were analyzed through the impedance plane to get the test results. However, these two signal parameters are sensitive not only to the defects, but also the heterogeneity of the material, sample size and sample structure, preventing the determination of the defect. Multi-frequency eddy current technology overcomes some shortcomings of single-frequency testing, but it can only provide limited data, the inconsecutive frequency of excitation will lost potentially useful information, so precise evaluation of the material can not be achieved, and it is difficult to form of intuitive achieving a visual defect.

Compared to conventional eddy current testing technology, pulsed eddy current testing technology is in the use of a repetitive broadband pulse (for example, square wave) excitation coil, the instantaneous current in the coil induced the transient eddy current in the sample, and with the rapid decay of the magnetic pulse together spread in the

material. Pulse signal contains a very wide spectrum, and a pulsed eddy current response is a continuous signal, the pulsed eddy current can provide a range of continuous multi-frequency excitation, so that the information of the response signal includes the different depth information. It is possible for quantitative evaluation of material. In addition, because it also runs a different current frequency, pulsed eddy current signal response is faster than the multi-frequency eddy current signals<sup>[4]</sup>. Therefore, the pulsed eddy current testing is an effective quantitative method for testing of metallic material.

### III. PEC EXPERIMENTAL SETUP

Fig.1 shows the PEC scanning system. The CNC-scanning machine is controlled by parallel interface. The QinetiQ TRESCAN® system is used for PEC measurements. The data acquisition setup is shown in Fig.2. A single period of an excitation waveform is created in Matlab and converted to an analog voltage signal by the Analog Output (AO) subsystem of a DAQ board (NI PCI-6255). The voltage signal is converted to the excitation current by TRESCAN box.

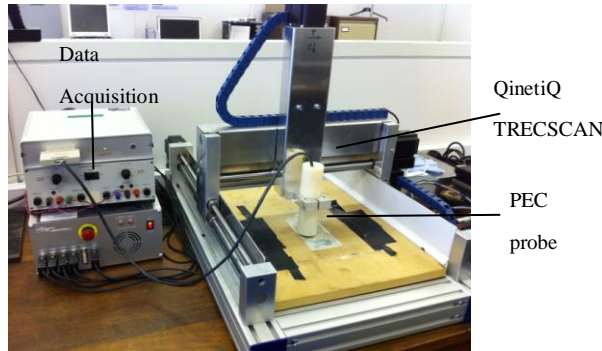


Figure 1. PEC set-up system

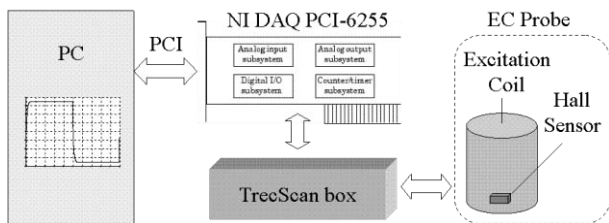
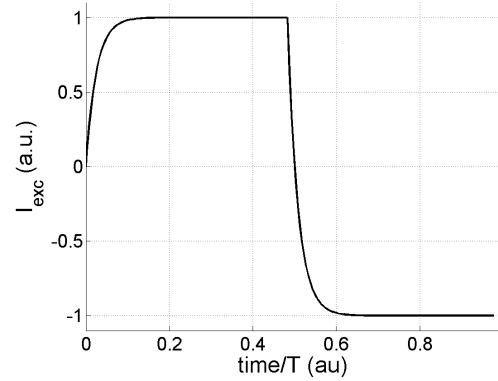


Figure 2. PEC acquisition scheme

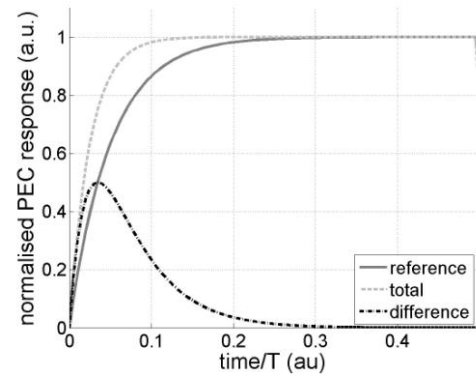
TRESCAN® operates in current excitation mode with an exponentially damped square wave of duty cycle 50% and time constant  $T_c = 100\mu\text{Sec}$ . The excitation current is fed to the excitation coil of the PEC probe. Hall sensor measures PEC probe response which is low-pass filtered (1kHz cut-off) and amplified by TRESCAN box. This signal is digitised by the Analog Input (AI) subsystem of the DAQ board. The acquired digital waveform is conditioned and analysed in Matlab. A self-shielded circular probe used in this work comprises a ferrite core. Excitation coil is 11mm in diameter. A Hall detector is the sensing element. The excitation repetition frequency is 200 Hz. The PEC response is acquired using a sampling rate of 500 kb/s.

### IV. PEC RESPONSE

Pulsed eddy current response signal has a very rich time-frequency characteristics, the researchers showed that in the defect detection of ferromagnetic material, the differential signal of the peak time, peak size, rising point and zero-crossing point can identify and characterize the defect information well<sup>[3][5]</sup>. In this paper, the peak differential signal is used to detect the information of notes.



(a) One period of PEC excitation



(b) One half of period of the normalised PEC response  
Figure 3. PEC response

Fig.3 (a) shows the input excitation signal for the sensor, the input is a square wave, showing a cycle of waveform. Fig.3 (b) shows the total normalized signal  $B$ , the reference signal  $B_{\text{REF}}$  and the differential signal  $\Delta B_{\text{norm}}$ .

The security line of notes actually contains both ferrous and non-ferrous material. In the use of pulsed eddy current technique for non-ferromagnetic material, the pulsed eddy current amplitude is affected weakly by the relative magnetic permeability, and when testing ferromagnetic material, the pulsed eddy current amplitude is affected by the magnetic effect<sup>[6]</sup> and magnetic characteristics. With the change of permeability it also changes obviously. In order to make the feature unified and effective, taking into account the characteristics of pulsed eddy current testing, this paper uses a signal processing method to eliminate the impact of magnetic permeability on the signal characteristics, relying on electrical conductivity to determine the characteristics of the security lines in notes.

Defining the maximum value of  $\Delta B$  to describe the characteristics of deviation of permeability, which is called  $\text{Max}(\Delta B)$ . In order to get the signal which changes due to electrical conductivity, variation in the signals were



normalised with maximum values of  $\Delta B$  before obtaining the difference signal. In this experiment, we take the signal from paper act as reference signal and the signal of security line as the total signal. Fig.3 (b) shows it can normalise the total signal  $B$  and reference signal  $B_{REF}$  to their respective maxima and to get the difference non-normalised signal  $\Delta B_{norm} = B/\max(B) - B_{REF}/\max(B_{REF})$ , where  $B/\max(B)$  is the normalised total signal and  $B_{REF}/\max(B_{REF})$  is the normalised reference signal. Due to the transient nature of the PEC diffusion in the material,  $\Delta B_{norm}$  peak will not be affected after processing<sup>[7]</sup>; at the same time, considering the use of sensors signal to extract feature, the useful signal may be overwhelmed by external electromagnetic interference<sup>[8]</sup>. In order to maximize signal resolution, we take the peak of  $\Delta B_{norm}$  to do follow-up analysis and 2D imaging.

## V. EXPERIMENTAL RESULTS ANALYSIS

In the experiment, a 20 pound, 5 pound notes and a piece of white paper were used, in accordance with the above parameter to set experiment.



Figure 4. Photo of notes

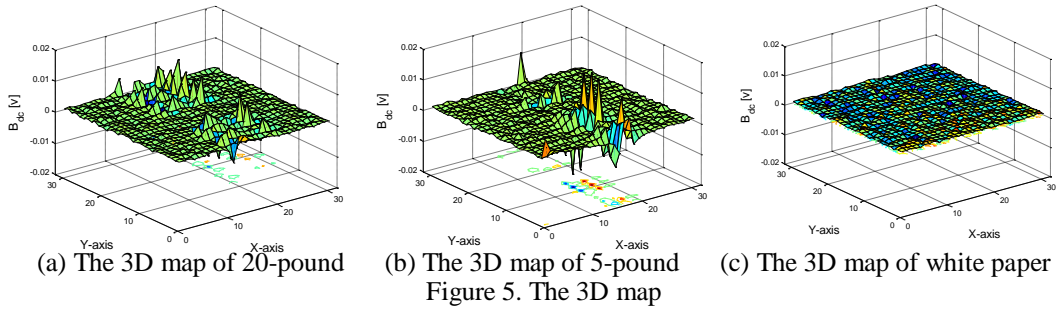


Figure 5. The 3D map

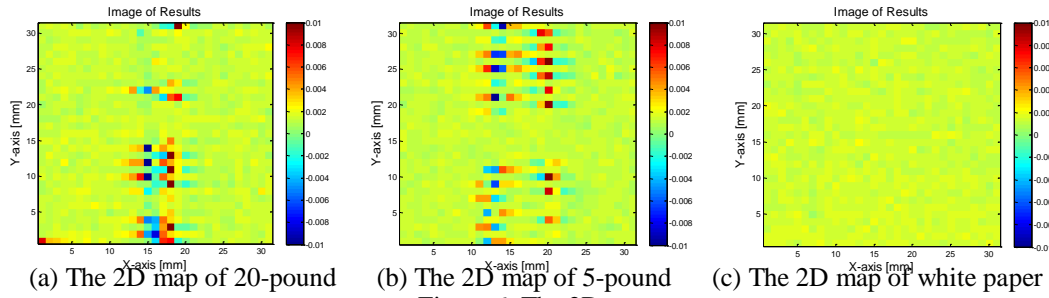


Figure 6. The 2D map

In the experiment, scanning on areas of notes where the security line exist. The scanning area was 30mm\*30mm, in X direction and Y direction, the scan step was 1mm. Get a total of 31\*31 scans points.

From Fig.4, it can be seen that the difference of security lines between 20-pound and 5-pound notes is not large, security line distributions of two notes are very similar. When the scan probe moves over the notes, the feature of the image has great difference. As shown in Fig.5 and Fig.6, the response signal of security lines and security line distribution has discontinuous area, mainly because the security line of notes is a "window style" layout. The embedded part and the exposed part of security line are alternative distribution. When the probe moved from the exposed part to the embedded part, the material covering in the security line caused the sensor lift-off effect, reduced the response signal.

From Fig.6, it can be seen that the response signal distribution is much wider than security line actually is. The signal distribution width of 20 pounds is about 20 mm and signal distribution width of 5 pound is only about 10 mm. On one hand, this is due to a certain size of the Hall element,

when the probe through the notes, the signals will have some expansion of coverage; on the other hand, the exposed part of security line in 20 pound is 5mm and the exposed part of security line in 5 pound is 4mm, the longer exposed part of security line makes a wider range of eddy current area, which generates differences between the different notes.

From Fig.5 it can be seen, when the probe passes through the notes and the paper, the 3D images are obviously different. In 3D images of notes, the peak value of eddy current is evident. In the 3D map of 20 pound and 5 pound, along the Y axis, the peak value of eddy current is throughout the plane. In the image of white paper, there is only a small disturbance and no signals. Compare the 2D image of notes with that of paper in Fig.6, there are significant differences in the color of images.

Therefore, the detection based on pulsed eddy current can effectively distinguish the authenticity of notes, while it can also effectively distinguish the various denominations based on their imaging characteristics.

## VI. CONCLUSIONS

In today's society, the counterfeit manufacturing technology is constantly updating, and for the identification of counterfeit technology, it is also constantly improving. So a number of specific test methods were proposed. By introducing the Pulsed Eddy Current technology for the detection of counterfeit notes, it can quantitatively analyze the security line of notes by pulsed eddy current detection. In this experiment, by analyzing the security line of notes quantitatively, we can see, the pulsed eddy current signals have a wide frequency, which contains rich amount of information. The signals also give high reliability of experiment and the experimental results are significant. At the same time, the requirement of the experiment is low and experimental operability is high. The results lay the foundation to the further analysis of the signal and it is expected to develop detectors which can distinguish between different denominations by further practical design and may count them at the same time.

Thank to School of Electrical, Electronic and Computer Engineering, NDT laboratories, Newcastle University for supporting the equipment and the experimental conditions.

#### REFERENCES

- [1] Jin Long Zhang, Identifying False Coin Instrument Based On Microcomputer[J]. Electronic Technology, vol.29, No.3, pp. 19-20, 2002. (In Chinese)
- [2] Qi Pang, Identification of Counterfeit Currency With laser technology[J]. Laser Technology & Applications, No.10, pp.9-10, 2005. (In Chinese)
- [3] Li Pingping, Zhou Guixiang, Realization of Measurement of Magnetic based on 80C196KC[J]. Machinery Design & Manufacture, No.5, pp. 147-149, 2006. (In Chinese)
- [4] Zhou Deqiang, Zhang Binqiang, Tian Guiyun, et al, Quantification of depth and classification of cracks using pulsed eddy current test technology[J]. Chinese Journal of Scientific Instrument, vol.30, No.6, pp.1190-1194, 2009. (In Chinese)
- [5] S.D.Karunanayaka, C.P.Gooneratne, S.C.Mukhopadhyay, et al, A Planar Electromagnetic Sensors Aided Non-destructive Testing of Currency Coins[J]. NDT, vol. 11, No.10, 2006.
- [6] Y.Z. He, F.L. Luo, M.C. Pan, et al, Defect classification based on rectangular pulsed eddy current sensor in different directions[J]. Sensors and Actuators A, vol.157, pp. 26-31, 2010.
- [7] Maxim Morozov, Gui Yun Tian, Philip J. Withers, The pulsed eddy current response to applied loading of various aluminium alloys[J]. NDT&E International vol.43, pp.493-500, 2010.
- [8] Maxim Morozov, John W Wilson, Gui Yun Tian, Measurement of Residual Stress in Mild Steel by Barkhausen Noise and Pulsed Eddy Currents[C]. 47th Annual Conference of the British Institute for Non-Destructive Testing, vol.50, pp.651-657, 2008.



# Infrared Thermography Study of Thermal Plume

Jafar Ali, Abdullah Abuhabaya and John Fieldhouse

School of Computing and Engineering  
University of Huddersfield  
Huddersfield, UK  
[j.ali@hud.ac.uk](mailto:j.ali@hud.ac.uk)

**Abstract**— This Study uses a thermography technique to determine the heat diffusion profile of thermal water discharge into rivers and canals using thermal imaging camera. It's provided thermal images for the surface of receiving water showing clearly the mixing zone, shape and edge of the plume which have been difficult to predict by mathematical models. The technique is one of the nondestructive testing tools which are predicting the diffusion of the heat without disturbing the fluid flow and its direction. The process applied on number of real canal sites and in laboratory to observe the actual thermal discharge. To verify the accuracy of the thermal images, the obtained data are compared with temperature measured by thermocouples on canal sites. The centerline temperature decay obtained from thermal images agreed with temperature measured by the thermocouples. This paper will be of interest to practicing engineers who deal with the environment.

**Keywords**- heat diffusion, plume, pollution, thermal image

## I. INTRODUCTION

The protection of the environment and natural resources from contamination by domestic and industrial wastes has been one of the major concerns of the general public and many environmental protection agencies for many years. The risk of thermal pollution increased in the recent years as the global warming began to threaten the universe. Therefore it is necessary that the scientists and involved organizations to develop continuance methods to predict waste heat distribution in the environment. In addition the climate change concerns are guiding most industries and commercial properties towards addressing their energy usage. In large buildings, where air conditioning is required, there is generally a need for "chillers" to control the temperature of the building. This process is not environmentally friendly and expensive in terms of energy used and maintenance issues. The alternative is to cool the buildings using natural resources such as induced wind drafts and water extraction from rivers and canals as the source of cooling. The system is abstracting cold water from the source to the heat exchangers and then discharge warm water back to it. The cooling water may be considered as thermal pollution if it is affected negatively on the chemical and biological balance of the receiving water. Therefore the behavior of the thermal plume in the receiving water must be determined. The studies of cooling water discharge into environment are many; but the mathematical approaches were unrealistic and extremely conservative

in their analysis so causing many valid proposals to be rejected. The current study is aimed at addressing that situation to provide a valid interactive analysis procedure that is better evaluate the potential of using any canal site for cooling purposes. The analysis concentrates on the use of thermal imaging camera to determine the heat diffusion profile of thermal plume discharged into the natural receiving water. The studies are performed on number of real waterways canal site and in laboratory experimental model tank.

The objective of this study is to investigate the behavior of the thermal discharge into the canal and the subsequent rise in water temperature. The outcome will be an evaluation formulation that will more accurately represent the discharge characteristics and additionally provide design guidance to allow marginal proposals to become viable.

## II. ENVIRONMENTAL CONSIDERATION

Water is usually withdrawn from canal to the heat exchanger of the cooling systems and returned to it after used; the returned water has a temperature greater than the ambient temperature of the canal which may cause the rise in temperature of the latter particularly in the region close to the outfall. The temperature of receiving water is an important factor to form the aquatic organisms since it is affected on the oxygen value, the foods and growth rate of the aquatic. the rise in the canal temperature in some regions may help the fish to flee from the hottest region to another which is giving those aquatic an activity helps them in growing [1]. Although, the water temperature is an important influence in aqua physiology, distribution and lifestyle, the increased of temperature above a certain degree (28°C in BW canal) jeopardize the aquatic life [2]. The dissolved oxygen and the solubility rate of oxygen will reduce when the canal ambient temperature exceed the 28°C, the reduction may reach such a level which fish and the other aquatics cannot survive. In addition the changes in the bulk temperature of canal raise the chemical and biological balance of the canal, and then reducing the quality of the water. Therefore it's important the efforts to carry out to control the temperature of the canal to provide a best condition for the aquatics.

## III. CANAL SITE

Numerous educational and business organizations utilize large office and computer facilities which generate heat, to prevent both occupants and their equipment from

overheating an environmental method of office and computer room cooling is required. Traditional air conditioning makes use of chillers to cool the system but if the building is located in close proximity to a natural water sources such as canals “Fig. 1”, then an environmentally friendly alternative may be available. “Fig. 1.b” shows a heated water discharge (plume) from Central Services Building at University of Huddersfield into Huddersfield Broad canal, “Fig. 1.c”. Cooling water is extracted direct from the canal and is pumped around a bank of plate heat exchangers and then discharged back into the canal. The area around the outfall called mixing zone and the moving turbulent water within the mixing

zone called plume, see “Fig. 1.b”. It is shown a very distinct velocity plume created by the semi-submerged discharge pipe causing turbulent flow and surface disruption. “Fig. 1.d” illustrate a typical canal site in which the discharge pipe is submerged and the discharge plume cannot be seen as clear as surface discharge.

In thermal discharge the majority of heated water moves to the surface of receiving water because the effects of buoyancy, whilst a part of heat transferred by convection to the layers below the discharge pipe. Therefore the evaluation of temperature distribution on surface of mixing zone will show the effects of canal ambient temperature by thermal discharge.

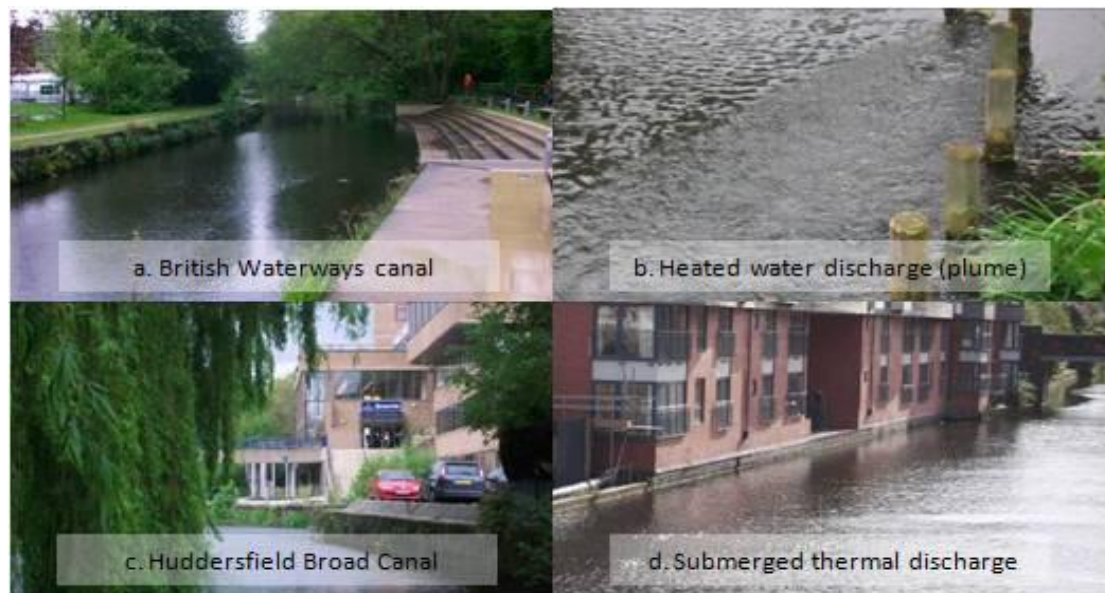


Figure 1: Typical British Waterways canal sites

#### IV. THERMAL IMAGES

Thermal imaging camera is a tool use to show the diffusion of heat on the surface of hot body. Human eyes are able to detect a visible light and can see a very small part of the electromagnetic spectrum. Infrared radiation is come from heat or thermal radiation such as sunlight, fire, radiators etc. lies between the visible lights and microwave portions of the electromagnetic spectrum which cannot be seen by our eyes. Thermal camera infrared thermography is transformed an infrared image coming from a hot body into radiometric translating the data into a colored image which is representative of the thermal gradients across the body [3]. The image may be observed on a LCD monitor and stored for future analysis and interrogation. The camera should set to the right emissivity of the test material; emissivity is the capacity of a surface to emit heat, at a given temperature, it is a relative quantity, which value varies from (0 to 1) for water is 0.96 [4]. To avoid unnecessary and unwanted reflections it is necessary to ensure the thermal imaging camera is placed in a position that is directed to the test body without the effects of reflection - which is the most problematic issue facing the thermographer - especially with thermal water studies. The effects of the reflections

may reduce if the thermal camera installed properly by directing the camera in an angle (angle on interrogation) of  $30^\circ$  with the surface of the receiving water. In some cases thermocouples may use to measure the actual temperature at the surface, then moving the camera to a viewing angle that measure the temperature similar to that measured by the thermocouple.

To indicate the mixing zone and temperature gradient within the discharge plume images were taken of the discharge at CSB site using a thermal imaging camera. “Fig. 2 and 3” show the comparison of the discharge plume at the Central Services Building (CSB) taken with the thermal image camera and a digital camera. The outlet pipe is semi-submerged and the plume can be seen on the canal surface, “Fig. 2” shows the extent of the plume and the differing temperatures of the timber protective posts around the discharge pipe. The variation of color on the timber posts especially on the top of the posts is due to reflection. “Fig. 2” shows plume and mixing zone from downstream. Thermal images can be obtained in a different group of colors; “Fig. 3” shows the plume with brighter colors and maximum temperature  $22^\circ\text{C}$ . The left hand side images in the figures represent thermal images whilst the right hand side photos are taken by normal digital camera.



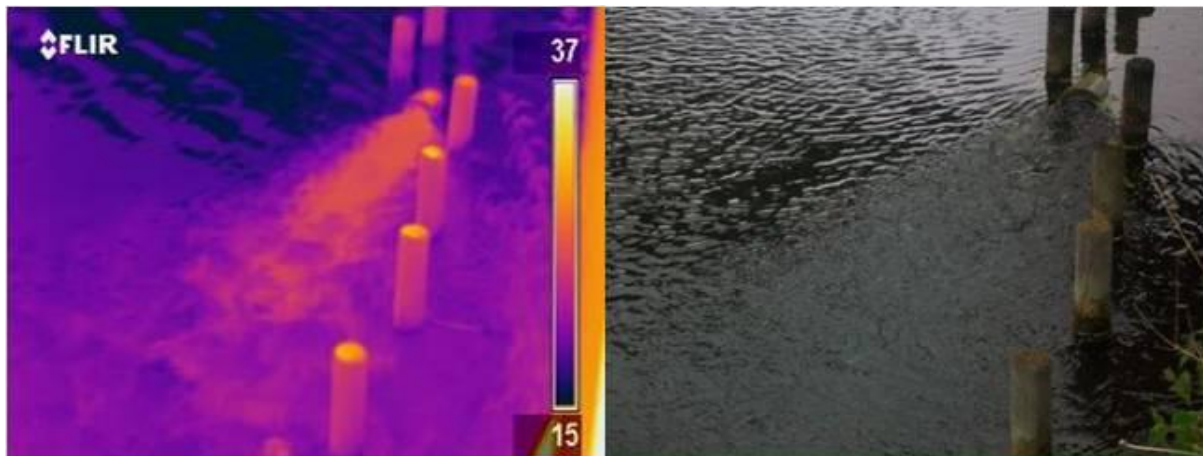


Figure 2: Thermal and digital image of thermal plume from downstream

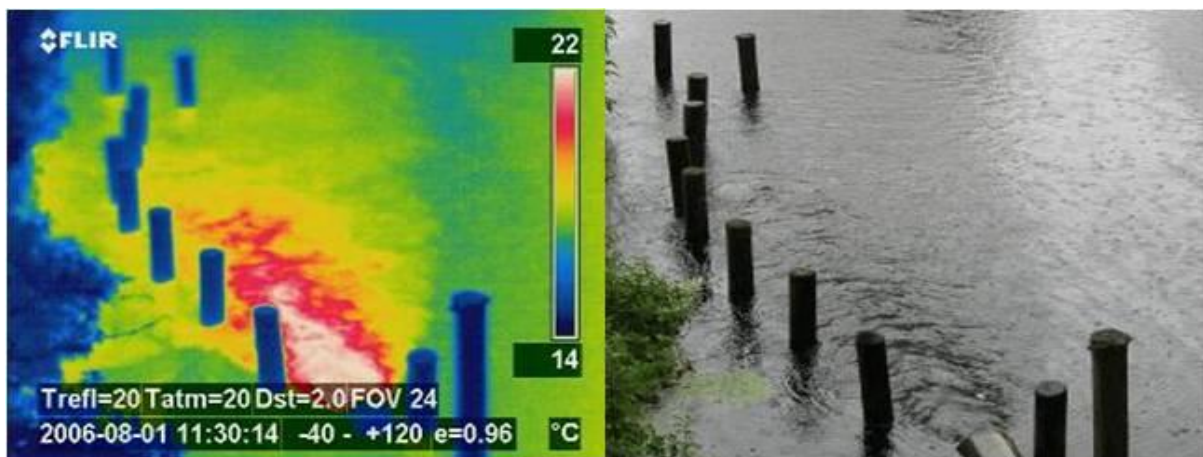


Figure 3: Thermal image in a different color with digital image

The previous images were for the thermal discharge on the surface of the canal. In the following figures thermal images will present for canal sites where the discharge pipe is deeply submerged. “Fig. 4” illustrates the end of the mixing zone on the surface of canal and edge of plume, the red color is the reflection of the opposite buildings as it appears in the digital photo for

the site on the right hand side of the figure. Thermal image in “Fig. 5” shows the edge of the mixing zone which is difficult to observe by normal digital camera except the area when the heated water reaches the surface. What can be seen also from the thermal image is the turbulence flow of the discharge plume.



Figure 4: Thermal and digital image for a deeply submerged discharge

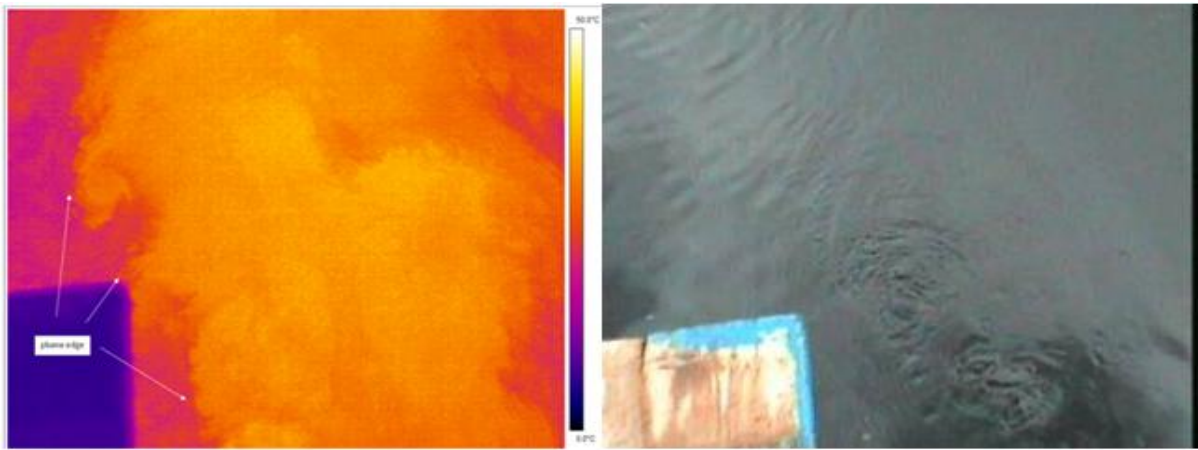


Figure 5: Mixing zone surface in submerged discharge

“Fig. 6” demonstrates the images for another canal site where the discharge pipe is submerged. The thermal image on the left hand side of the figure shows green spots which is the area where the plume reaches the

surface, whereas the red color is due to reflection on the embankment. Not always the thermal plume reaches the free surface of canal because the temperature balance achieved in the layers below the surface.

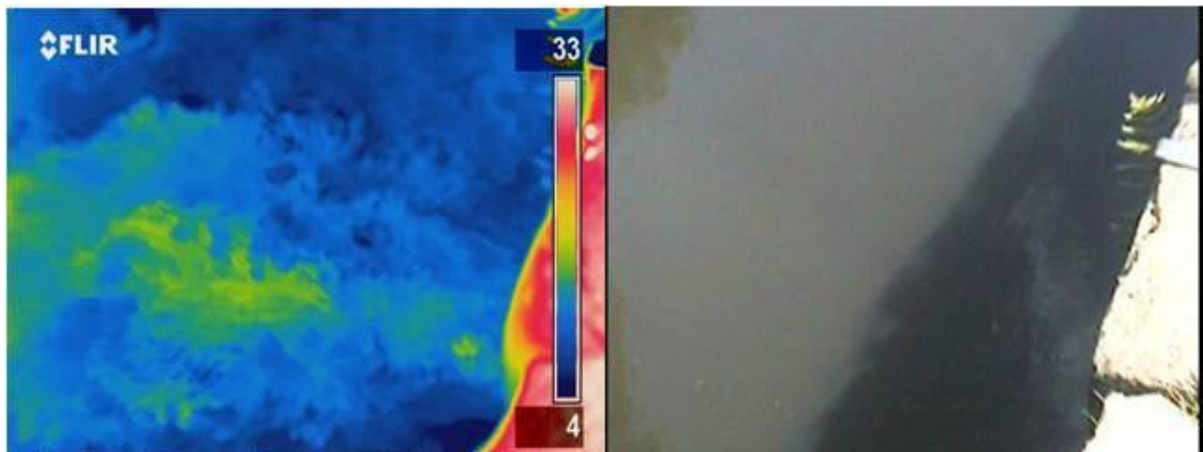


Figure 6: Images of submerged thermal plume when reaches the surface of canal

## V. TEMPERATURE MEASUREMENTS

To verify the temperature measured by thermal imaging camera, a survey carried out on the canal site by using thermocouples to measure temperature of the plume. Initial temperature measurements were taken on particularly hot bright sunny days “20°C” ambient for the CSB site at University of Huddersfield. To monitor the whole site it was necessary to establish a reference grid over which the results of water flow and temperature would be measured across, along and below the surface of the canal. The centerline of the outlet pipe and the end position of the pipe served as the zero position. A graduated pole was laid along the banking to give the linear reference points and a second pole was fixed at 90° to the banking to give the transverse reference positions. A third vertical pole was secured to the main pole to which the thermocouple probe and flow meter could be attached. This third pole was also graduated to indicate depth. The arrangement is generally as shown in “Fig. 7”.

Ambient canal water and air temperatures were recorded with the use of ‘k’ type thermocouples and digital meter at the start and end of the trial [5]. The canal water temperature was then measured and recorded at the grid points.



Figure 7: Pole holds thermocouples probe and flow meter

The measurements taken at the centerline of plume 50mm below the surface indicate the maximum core temperature of the plume is contained within an area 400mm wide by 1meter long equally displaced about the centerline of the discharge pipe, this core is not evident at the mid-depth indicating the turbulent flow possibly occurs close to the diameter of the discharge pipe. “Fig. 9” shows the relative distribution of temperature at the centerline of plume along the 5m length of the plume on the embankment side.

Thermal imaging camera is used widely in other applications such as electrical, mechanical, civil and chemical engineering. It is used by the maintenance engineers and inspectors to find the cable cuts, water leak in pipes, heat pump applications, boilers, heat exchangers, brake disc studies and so on.

## VI. MATHEMATICAL ANALYSIS

There are many mathematical investigations available on thermal discharge studies. One of the wider studies on the surface discharge is carried out by [6]. Their study reviewed the previous work on thermal discharge up to 1978, They solve nine partial differential equations using finite difference, the continuity equation, two momentum equations in x and y direction, thermal energy equation, state of water, the kinetic energy equations and two equations for turbulence. The work as any other mathematical analysis started with lot of assumptions. The surface thermal discharge continually been studied by the researchers, an integral mathematical model by [7] is presented but the number of the assumptions in his work are more than those in the numerical analysis. A comprehensive study on thermal submerged discharge is undertaken by [8], this study modeled the plume property below the free surface of the receiving water.

In the current paper the diffusion of the thermal plume on the free surface of the receiving water has been predicted without complicated mathematical analysis or software. What can't be seen in the current work is the behavior of the thermal plume below the surface, although the thermal plume is deflecting to the surface. To fulfill this gap a simple mathematical model is produced to predict the plume behavior below the free surface of the receiving water. The model derived from the heat advection diffusion equation “1”:

$$\frac{\partial T}{\partial t} + U \frac{\partial T}{\partial x} + V \frac{\partial T}{\partial y} + W \frac{\partial T}{\partial z} = D_x \frac{\partial^2 T}{\partial x^2} + D_y \frac{\partial^2 T}{\partial y^2} + D_z \frac{\partial^2 T}{\partial z^2} \quad (1)$$

Where T is the temperature, U, V and W are velocities in x, y and z directions respectively,  $D_x$ ,  $D_y$  and  $D_z$  are turbulent diffusivity in x, y and z directions. The procedure followed to solve the partial differential equation “1” by [9], and the following equation “2” is produced [5]:

$$T(x, z) = \left( \frac{T_0 - T_a}{2} \right) \left( \operatorname{erf} \frac{b-z}{2\sqrt{p \cdot x}} + \operatorname{erf} \frac{b+z}{2\sqrt{p \cdot x}} \right) + T_a$$

Where  $T_0$  and  $T_a$  is discharge and ambient temperature respectively, b is discharge pipe diameter and “ $p = D_z/U$ ”. The vertical heat diffusion predicts by “1” is illustrated in “Fig. 8”; it is shown the diffusion of the thermal plume vertically below the surface. The vertical diffusion is small in comparison with the lateral diffusion. Although the simplicity of the “2”, it gives a good prediction for the thermal plume heat diffusion below the free surface.

## VII. RESULTS AND DISCUSSIONS

The existing mathematical models on thermal discharge are modeling mainly the temperature decay along or across the centerline of the plume. Some of the studies predicted the surface temperature, whereas the edge of the plume and boundary layers were poorly defined. In addition all the studies are required initial assumption to simplify the equations of flow and heat transfer, which may criticize the results. The current study on thermal discharge is presenting with very high definition the heat diffusion profile which most of the previous studies failed to achieve. The thermal images show the exact surface area of the mixing zone and plume. The edge of the plume and boundary layers can be clearly identified. “Fig. 9” shows the comparison of the plume centerline temperature measured by the thermocouple and the thermal images. The centerline temperature decay obtained from the thermal images agreed with the data obtained by thermocouples. The slight difference in both results because the thermal images data are for the free surface whilst the thermocouple data are for a layer 50mm below the surface. For the submerged discharge, the thermal plume effect is not observed until the warm water rises to the surface.

Limitation of the study is that the thermal imaging camera will not be able to predict the temperature distribution below the free surface of receiving water. However the thermal plume is a buoyant discharge and subject to buoyancy effects which may raises the plume to surface within a certain distance from the outfall.

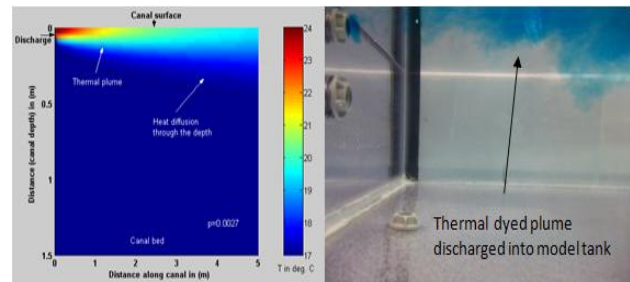


Figure 8: Heat diffusion below the free surface for the mathematical model and comparison with the laboratory experimental dyed plume



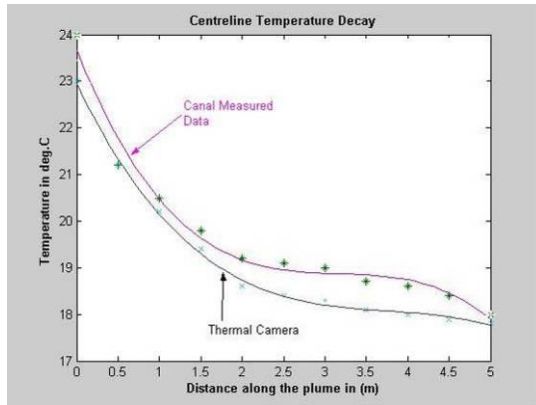


Figure 9: Temperature decay along the centerline of the plume measured by thermal imaging camera and comparison with the thermocouples measured data

A simplify mathematical model is developed to predict the behavior of the plume below the surface. The effects of the emissivity on the thermal images are demonstrated in “Fig. 10”.

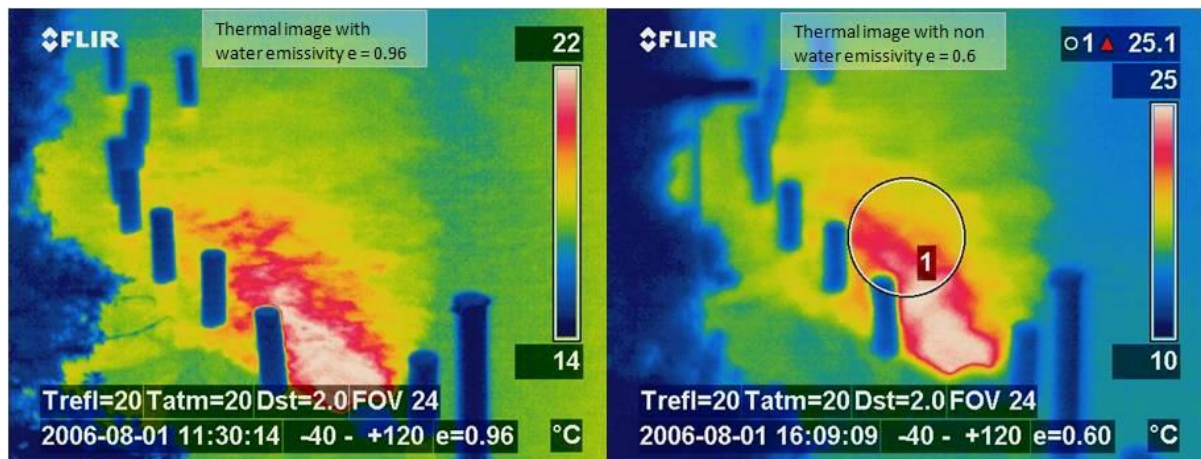


Figure 10: Thermal image of the plume using water emissivity 0.96 on the left and non water emissivity 0.6 on the right.

## REFERENCES

- [1] C. Hoar, “Fish response to discharge events from a power plant cooling reservoir in a river affected by acid mine drainage and thermal influences,” M.Sc. Dissertation, West Virginia University, USA, 2005.
- [2] Environment Agency, The Surface Water (Fishlife) (Classification) Regulations No. 1331, London, 1997.
- [3] Thermal CAM Researcher, FLIR Systems Users Manual, Professional Edition, Version 2.7 SR-1
- [4] J. Coulson and J. Richardson, Chemical Engineering, 6th ed. vol.1, Butterworth Heinemann, 1999.
- [5] J. Ali, J. Fieldhouse, C. Talbot and R. Mishra, “Thermal discharge of warm water into cooler stagnant water,” Int. Symp. What Where When Multi-dimensional Advances for Industrial Process Monitoring, Leeds, pp. 23 – 24, 2009.
- [6] J. McGuirk, and W. Rodi, “Mathematical modelling of three-dimensional heated surface jets”, Journal of Fluid Mechanics, 95, (4), 1978, pp. 609-633.
- [7] A. Zaghloul, “Turbulence Structure of Plane Surface Jets in a Current”, Ph.D. thesis, University of Western Ontario, Canada, 1996.
- [8] G. Jirka, “Integral Model for Turbulent Buoyant Jets in Unbounded Stratified Flows”, Environmental Fluid Mechanics, 4, 2004, pp.1-56.
- [9] J. Crank, “The Mathematics of Diffusion”, Oxford.19 853307 1, 1970

The image on the left has been taken with the correct emissivity for water, whilst the one on the right has been taken with an emissivity of “0.6”. Therefore the first appears clearer than the latter.

## VIII. CONCLUSION

The mixing zone, temperature distribution within the surface of the receiving water and the shape of plume are clearly defined by the use of the thermal imaging camera without any mathematical assumption. The temperature distribution obtained by thermocouples verified the temperature distribution of the discharge plume obtained using the thermal image camera. The thermal images are influenced by the emissivity of the tested body and reflections. To avoid these problems, it's necessary to ensure the camera is set to the right emissivity (0.96 for water) and placed in a position that is directed to the tested body without effects of reflection. Further investigation may be required to find the temperature distribution through the depth of canal and the thickness of the plume.



# A Wireless Sensor Network based Structural Health Monitoring System for an Airplane

Jasleen. K. Notay, Ghazanfar. A. Safdar  
Department of Computer Science and Technology  
University of Bedfordshire

Park Square, Luton, LU1 3JU, UK  
{jasleen.notay@study.beds.ac.uk, ghazanfar.safdar@beds.ac.uk}

**Abstract**— Structural Health Monitoring (SHM) is a mechanism that is used to determine the origin of any damage in a particular structure and to evaluate the health of civil structures and buildings. In traditional SHM systems, usually, sensors are embedded into a structure and any information sensed from the vibrations of the structure is sent via wires to a central data repository. Such systems suffer from drawbacks like expensive installation and maintenance, unreliable communication in wires etc. Wires in modern SHM systems are being replaced by Wireless Sensor Networks (WSN) that alleviate many disadvantages in traditional wired SHM systems. In this paper, a WSN based SHM system is designed for an airplane with sensors placed at certain optimum locations inside it. The WSN based SHM model for airplane is tested for its performance using OPNET and it is found that the proposed model functions efficiently in a simulated environment.

**Keywords**- *Wireless Sensor Networks, Structural Health Monitoring, Airplane.*

## I. INTRODUCTION

All structural entities are exposed to aging, disintegration and wear and tear due to natural or catastrophic reasons [1]. Structural Health Monitoring (SHM) is a technique that is used to analyze and evaluate the stability and health of a structure; the functional lifetime of the structure can also be calculated on the basis of its existing state [2]. SHM techniques have been employed in bridges [3], aircrafts [4], trains, historical buildings [5] and so on and sensors are the main components that make up a SHM system. According to [6], traditional SHM systems bear comparison with laboratory based data acquisition units wherein each sensor is connected to a central unit using a co-axial cable. Sensed data is converted to digital form before being sent to the central unit for analysis by monitoring algorithms. Contemporary SHM systems suffer from many drawbacks mainly because of the long wires that need to be used to connect each sensor to a central unit. Large sized structures call for installation of longer wires which is both time consuming and labor intensive; for example, deployment of a medium sized SHM accounts for 75% of the total testing time and about 25% of the installation costs [6]. High installation costs may dramatically reduce the number of sensors that can be made part of the SHM system [7] and limited number of sensors cannot provide accurate information about a large

structure. Moreover, wires are vulnerable to breakage and in some cases may be chewed away by rodents [6]. In order to combat the numerous disadvantages of a wired SHM system; wired sensors are being replaced with Wireless Sensor Networks (WSNs). Wireless sensors being much less costly as compared to their wired counterparts are easy to install inside a structure; thus making the wireless SHM deployment cost effective and less labor intensive. In this way, more number of sensors can be made part of this wireless monitoring system to provide accurate information about the structure.

An airplane uses different types of sensors to monitor different parameters at different locations like calculating the level of fuel in the fuel tanks, detecting fires and smoke in different departments, measuring temperature and pressure inside a compartment or the tyres etc.; the health information sensed by each sensor is transmitted to the electronic system inside the airplane using a coaxial cable [8]. It is a well known fact that weight plays a crucial role in the operation of the airplane; more the weight of the plane, greater is the consumption of fuel. Thus, employing a WSN based SHM system for an airplane would result not only in the reduction in the number of wires and cables that need to be installed, but also lead to lesser fuel consumption because elimination of wires from certain areas inside the airplane would make it light in weight. Reduced fuel consumption would provide favorable environmental and economic advantages; airplanes would emit fewer harmful emissions into the environment and at the same time incur cost savings and be capable of allowing passengers to carry more weight per individual. Apart from the need for eliminating long wires and cables and reducing overall system weight and cost; another major advantage of a SHM system is the automation of maintenance routines and the advancement from schedule based maintenance to condition based maintenance [9]. The advantages of condition-based maintenance over schedule based maintenance include prompt maintenance techniques and routines, greater availability of air fleet and decreased total cost of ownership which otherwise is increased by unscheduled maintenance requirements [9]. SHM systems can be designed to not only detect damage, but also to identify the location of damage within the structure, calculate the extent of damage and the impact

that the damage would have on the structure for its remaining lifetime [9].

This paper proposes a WSN topology for SHM inside an airplane. The proposed SHM system would act like a nervous system for health monitoring of identified six vital locations inside the airplane. Each sensor would wirelessly communicate monitoring information into the indicating system of the airplane for the crew to view and take action accordingly. The proposed network is simulated in OPNET and the results are analyzed and interpreted. The paper is structured as follows: Section II is a discussion about the data bus and the indicating systems inside an airplane and their respective functions; section III highlights the locations inside the plane where sensors can possibly be embedded; the structure and function of the proposed model for SHM is presented in section IV; results obtained by simulating the network in OPNET are provided in section V and the conclusions are summed up in section VI.

## II. EXISTING WIRED INFRASTRUCTURE INSIDE AN AIRPLANE FOR SHM

Before we could discuss about our proposed WSN based SHM scheme, it is important to provide an insight into how various avionic units of the airplane function together as one system and the manner in which they transmit SHM related information amongst each other.

### A. The Data Bus inside an airplane:

Different avionic systems inside the airplane communicate with each other using data buses inside the airplane. All large passenger aircrafts use a serial data bus (only a single bit of data can be transmitted over the bus at a time) to transfer data to and from avionic units [8]. Many different kinds of serial data buses are present in an aircraft that carry different kinds of data.

### B. Display Units inside an aircraft:

Different kinds of indicating units inside the cockpit display the condition and performance of various avionic units on board for the crew to view and also indicate when a particular unit ceases to function. Major passenger aircrafts employ one of the following two kinds of indicating systems for the purpose, namely,

- The Engine Indication and Crew Alerting System (EICAS).
- The Electronic Centralized Aircraft Monitoring System (ECAM).

The function of both of the above mentioned systems is to audit aircraft systems and display relevant alerts and warnings on the indicating units inside the cockpit for the crew to take appropriate action. Sensors installed inside different systems sense health information like engine thrust, engine rotational speed, exhaust gas temperature, cargo bay temperature, tyre pressure etc. [8] and feed it via a wiring harness into the data bus of the airplane so

that it could be routed to the indicating system of the airplane for display.

## III. WSN BASED SHM SYSTEM FOR AN AIRPLANE

Having gone through the drawbacks of wire based SHM systems and the advantages of using WSNs (Section I), this section provides some of the prerequisites that have been identified for an SHM system based on WSN. Additionally, identification of some critical locations for the placement of WSN inside the airplane for SHM also forms part of this section.

- The wireless sensors used as part of the SHM system would be autonomous in nature [10].
- In order for the network to be scalable and robust; the topology should be hierarchical with the lowest tier being clusters.
- Since the areas to be monitored inside an airplane may be large in size, hence any communication between the cluster head and the sink/central data repository would be performed in a multi hop manner.
- Computation of data consumes lesser power consumption than communicating it (although this also depends on the transceiver and the microcontroller in use) [11] hence data aggregation mechanisms and in-network processing would be performed in order to reduce data redundancy and communication and power costs.

### A. Identification of critical locations inside an airplane:

Although there could be numerous locations, but in order to form a SHM system based on WSNs, it is suggested that sensors be deployed at certain optimum locations inside the airplane that need continuous monitoring, enumerated as follows:

#### 1) The Fuel Tank:

Passenger aircrafts have integral fuel tanks located in the wings and the tail of the airplane. Solid ribs divide each wing into sections such that two fuel tanks can be accommodated in each wing. The central portion between the two wings also serves as a fuel tank [8]. Sensors need to be placed inside each fuel tank to measure the level of fuel.

#### 2) Inside the exhaust:

Sensors placed inside the exhaust would indicate if any obstructions are present inside it.

#### 3) Aircraft Wheels:

Although the health and condition of the wheels is always checked before take-off and after landing; there may arise possibilities where wheels might get damaged while on the runway or in the air. Sensors placed inside the wheels would report wheel health post take-off and

pre-landing and also alert if some wear and tear appears in them or if there is enough air inside them.

#### 4) *Engine:*

The engine is the heart of an airplane and it is vital that it does not get overheated or physically damaged. Sensors installed in and around the engine would be tasked with monitoring the temperature of the engine surrounds and the state of its components.

#### 5) *Wings:*

Wings play a vital role in balancing the aircraft and are exposed to corrosion and aging; hence sensors installed in the wings would sense ambient vibrations arising from them and report about any cracks or damage caused by corrosion or any other source.

#### 6) *Fire and safety:*

Certain areas inside the plane for example the cargo decks and the passenger area carry combustible items like luggage and food items and fire in these areas can prove fatal. Smoke detectors or sensors in certain areas like the cockpit, the kitchen, passenger section and the cargo decks may pick up indications of a fire starting and send alerts through the SHM system.

### IV. PROPOSED TOPOLOGY

Having identified locations of deployment, sensors need to be placed according to a network topology so that sensors placed in different locations can communicate with each other effectively. This section describes a cluster based, three tiered WSN topology for SHM in an airplane as illustrated in Figure 1. The description of the various hierarchies is provided as under:

#### A. *The hierarchical model:*

##### 1) *Tier 1:*

The lowest level of the topology is formed by sensors of different kinds which are tasked with sensing activities that depend on the monitoring activity required from them; for example, smoke detectors would be deployed for fire and safety purposes, pressure sensors inside tyres, temperature sensors in engine surrounds etc. As highlighted in Section III above, the sensors should be autonomous in their function so that instead of reporting every sensed measurement to the microcontroller, the sensor could examine its own sensed data and report to the microcontroller only when an actual event occurs or a threshold is exceeded [10]. This would help in conserving power as the microcontroller accounts for maximum power consumption. Sensors are arranged in the form of clusters so that their limited resources (power and lifetime) could be combined in order that overall network lifetime could be increased. Each cluster has a leader acting as a cluster head [12].

Since sensor nodes would be either embedded deep inside the structure or installed in hard to reach places

inside the airplane, thus it is important to have some energy saving mechanism in place for the sensor nodes. Paying careful consideration to the design of the Media Access (MAC) layer would reduce the requirement of frequent recharging/replacing batteries of sensor nodes. Furthermore, intelligent scheduling by the cluster heads would enable sensors in a cluster to have prolonged doze mode and to become active only for periodic sensing or during the occurrence of an event.

##### 2) *Tier 2:*

Cluster heads of each cluster of sensors constitute the second tier of the model and their features are explained below.

##### a) *Cluster Heads:*

Given the vast size of certain parts of the airplane where health monitoring has to be performed (for example: wings, fuel tanks, cargo decks etc.), a large number of sensor nodes are required to provide accurate statistics of the sensed information. The best method of integrating the already limited resources of the sensor nodes is to organize them into clusters with a cluster head responsible for each cluster, thus the second tier is formed by cluster heads of the respective clusters. Clusters result in a stable, scalable and robust topology because new sensor nodes could easily be incorporated in the network while any damaged ones could be replaced and communication is limited between the sensor nodes and the cluster head; thus node energy is not wasted in communicating over long distances. It should be noted that the number of clusters in a particular area inside the airplane would vary with its size and the number of sensors deployed. Every cluster head in the topology is assigned the following responsibilities:

- Query and gather sensed measurements from the sensors.
- Aggregate collected information with the objective of eliminating redundancy, minimizing communication costs and summarizing voluminous data into more accurate information [13].

Cluster heads have very important functions to perform and should not stop working in case they run out of battery power, that is why, in this case, they are provided with a secondary source of energy from the power supply units inside the aircraft (which includes batteries, generators, inverters and transformers) [8].

##### 3) *Tier 3:*

The next level in the model comprises of wireless access points and a central data repository.

##### a) *Access Point:*

Sensors (in the form of clusters) would be placed very far away from each other on the airplane. For inter cluster communication to take place, access points are installed that facilitate interaction between different cluster heads and the central data repository (explained below). The

access point is externally powered up from power sources on the airplane.

*b) Central data repository:*

The central data repository used for this topology is synonymous to a sink in a WSN. It contains health monitoring algorithms that analyze aggregated measurements received from all cluster heads and draw conclusions about the current structural state accordingly. Like the cluster heads, the central data repository too is powered externally by the power sources on the airplane (Figure 1).

*B. Features and operation of the model:*

With sensors installed in the six locations as already explained in Section III and their cluster heads chosen, the WSN would operate in the Industrial, Scientific and Medical band (ISM) band since sensors are capable of communicating only over low data rates. Sensors in a cluster would be in direct communication with their cluster head and are programmed to sense in low duty cycles where they sense the structure for low-frequency changes as well as alert when an event like crossing of a set threshold occurs [14].

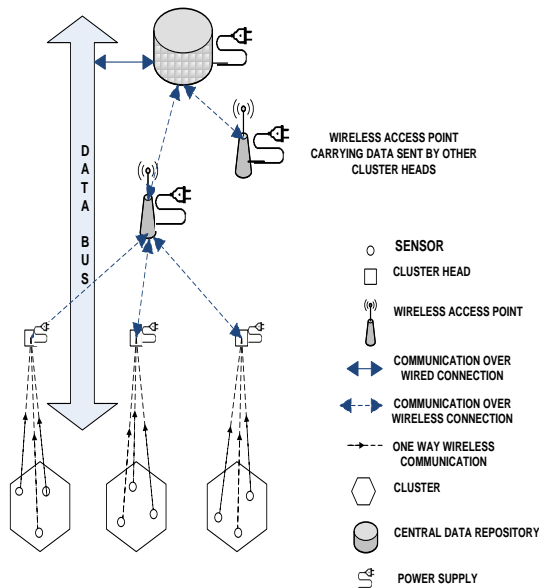


Figure 1: Proposed topology for SHM in airplanes

The cluster heads support two kinds of communication: inter cluster and intra cluster communication. Intra cluster communication occurs when cluster heads query sensors for sensed information and any data is received from the sensors. Inter cluster communication occurs when a cluster head serves as a relay node to route aggregated information from a different cluster head to the nearest access point. Inter and intra cluster communication takes place on different frequencies of the ISM band in order to avoid any interference between the two. In order that cluster heads do not get over worked with their tasks, it is suggested that backup cluster heads also be installed so that they could replace existing cluster heads in case of failure or any such situation.

All access points send the information that they receive from various cluster heads to a central data repository which contains health monitoring algorithms that analyze all sensed readings received from different parts of the airplane, filter it and conclude about the current state of the monitored structure, decide if an event has occurred, how severe the event is, what the remaining lifetime of the structure is and if any emergency indications need to be indicated etc. All conclusive information about each of the monitored structures (like engine, wings etc.) is put onto the data bus of the airplane and displayed in color coded format of the ECAM/EICAS system according to the state of urgency. Thus, any events detected, or thresholds exceeded or even the smooth operation of the particular structure is first analyzed by the central data repository and then fed for display on the ECAM/EICAS display units in the cockpit.

Figure 2 and 3 illustrate how the wireless network is deployed inside the fuel tank and on the wings of the plane.

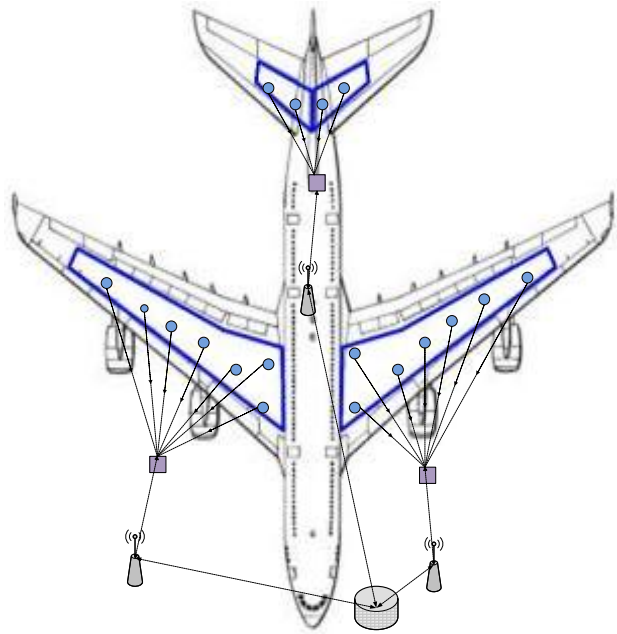


Figure 2: WSN installed inside the fuel tanks

*C. Compatibility between 802.11 based access point and 802.15.4 based sensor nodes*

The topology incorporates a wireless access point into a WSN; both the access point and sensor nodes belong to different standards that are incompatible with each other: 802.11 and 802.15.4 respectively. However, [15] discusses about how an 'Adaptation Layer' can be introduced in between the data link and the network layer of the 802.15.4 protocol stack to enable routing of IPv6 packets over an 802.15.4 network by employing header compression and fragmentation mechanisms to accommodate an IPv6 packet transmission in a low rate wireless sensor network. Thus, introducing an adaptation layer in the existing protocol stack for WSNs can support the operation of WLAN and WSN nodes in the same network.

## V. TESTING THE PERFORMANCE OF THE PROPOSED TOPOLOGY

The performance of the proposed network is simulated in OPNET Modeler 16 using the Zigbee model as a platform. The Zigbee model was chosen because OPNET does not provide a sensor model and because its features bear similarity to those of a wireless sensor network. The IEEE 802.15.4 standard provides the specification for the physical and MAC sub layer of Low Rate-Wireless Personal Area Networks (LR-WPAN) networks (both wireless sensor networks and the zigbee networks are examples of LR-WPAN networks) and both zigbee and wireless sensor networks share the same physical and MAC layers [7].

Three scenarios were constructed to simulate the communication of a cluster(s) with the access point via a cluster head(s). A zigbee coordinator was treated as an access point, a zigbee router was used as a cluster head and zigbee nodes performed the function of sensor nodes of the proposed topology.

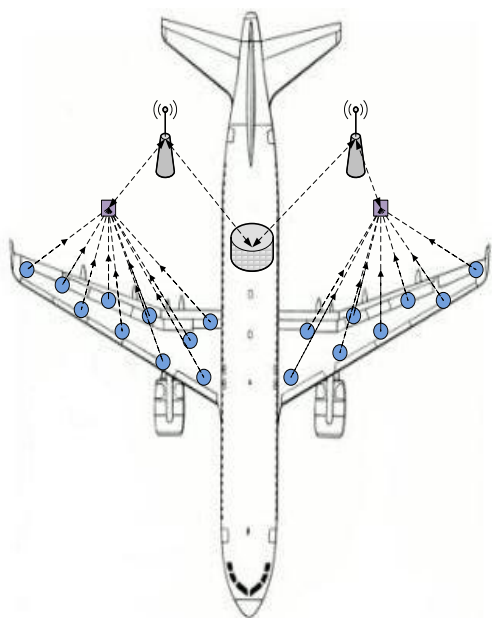


Figure 3: WSN deployed on wings

Scenario 1 consisted of two clusters each with a cluster head and three sensor nodes, communicating with an access point. Scenario 2 and 3, both comprised of one cluster of 3 sensor nodes communicating with one access point; the only difference being that nodes in scenario 3 were placed at a greater distance from the access point as compared to those in scenario 2. Variation of MAC parameters: data dropped, delay and throughput was observed with number of nodes and the distance between the nodes as the results of the simulation. Only network related performance is evaluated taking into account factors like throughput, dropped data packets etc. We believe that throughput is an important parameter which shows that the sensed information is actually being sent with certain percentage of data dropped, thus it proves that the sensor nodes are doing their job for which

WSN based SHM is proposed. The following conclusions were drawn:

Data dropped in the network is directly proportional to the number of sensor nodes. It dropped from 180 bits/sec for scenario 1 to 90 bit/sec for scenario 2 and 3. Reasons for such a pattern are that nodes do not sense all the time; they are in fact programmed to sense periodically for data and spend the rest of the time in sleep mode. Secondly, a network takes time to enter into the steady state (refer figure 4). In the transient state, a network always shows higher data dropped and lower throughput. As a consequence of data dropped, packets need to be retransmitted and hence delay experienced by the network as a whole also increases. Larger the number of nodes, greater would be the data dropped and hence greater would be the delay experienced by the packets in the network (see figure 5). It is also observed that although scenario 1 shows greater delay as compared to that of scenario 2 and 3; as the network progresses into the steady state, due to saturation of queues, delay becomes independent of the number of nodes and becomes nearly the same for all three scenarios.

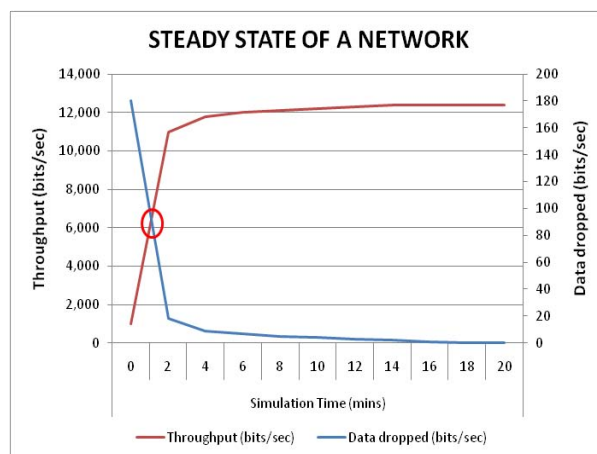


Figure 4: Throughput vs Data Dropped

Throughput of the network increases with the number of nodes (12,000 bits/sec for scenario 1 and 6,000 bits/sec for scenario 2 and 3) and decreased data dropped rate. Upon plotting the readings for throughput and data dropped on the same graph (Figure 4), we can infer that the proposed topology enters into steady state 1 minute after the network is powered on. In the beginning, the network functions in the transient state with high data dropped and low throughput while the results are vice versa when the network enters into steady state (because of reasons explained already). The variation of throughput and data dropped with the number of nodes for three different scenarios is shown in Figure 6.



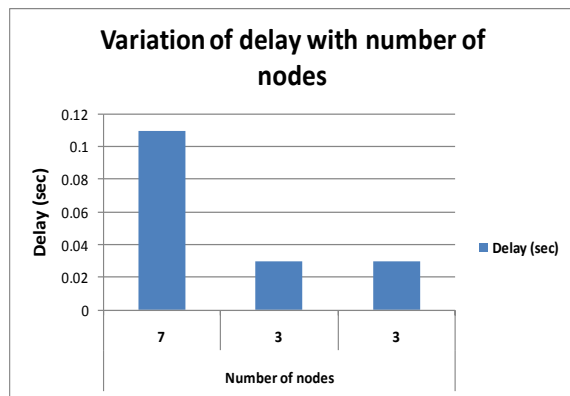


Figure 5: Delay experienced by the sensor nodes

It is important to note that network throughput parameter for scenario 2 and 3 showed no variation and was recorded to be the same which means that the performance is not affected by the distance between the nodes (as long as the nodes lie within communication range of each other).

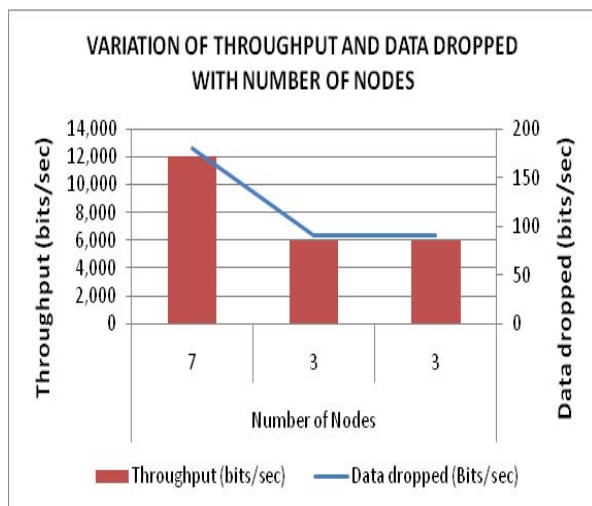


Figure 6: Variation of throughput and data dropped.

## VI. CONCLUSION AND FURTHER WORK

The work presented in this paper is in the stages of its infancy but it is promising. As part of this master studies research project, a WSN based topology for SHM in an airplane is proposed and is tested in OPNET modeler. It is concluded that the proposed topology does operate efficiently, however, the performance testing of the proposed network should be enriched by incorporating more parameters and some rigorous testing; a limitation was imposed on doing the same because OPNET does not provide a model for wireless sensor network. As part of future work, the proposed framework could be evaluated by practically building a prototype using WSN hardware /

software and doing some performance analysis before the actual deployment is made in a real world environment. The prototype could further involve setup of a central repository with added policies to show how sensed information is being processed and then actions taken accordingly.

## REFERENCES

- [1] Lynch, J.P. and Loh, K.J. (2006) 'A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring', *The Shock and Vibration Digest*, 38(2), pp. 91-128, SAGE Publications.
- [2] Chou, P. H. and Park, C. (2006) 'Energy-Efficient Platform Designs for Real-World Wireless Sensing Applications', *VLSI Design, Automation and Test, 2006 International Symposium on*, IEEE Computer Society.
- [3] Kim, S. et al. (2007) 'Health Monitoring of Civil Infrastructures Using Wireless Sensor Networks', *Proceedings of the 6th international conference on Information processing in sensor networks*. ACM Digital Library.
- [4] Speckmann, H and Roesner, H. (2006) 'Structural Health Monitoring: A Contribution to the Intelligent Aircraft Structure', *European Conference for Non-Destructive Testing*.
- [5] Anastasi, G., Lo Re, G. and Ortolani, M. (No Date) 'WSNs for Structural Health Monitoring of Historical Buildings'.
- [6] Kottapalli, V.A. et al. (2003) 'Two-Tiered Wireless Sensor Network Architecture for Structural Health Monitoring' *SPIE's 10<sup>th</sup> Annual International Symposium on Smart Structures and Materials (USA)*, SPIE Digital Library.
- [7] Koubâa, A., Alves, M. and Tovar, E. (2005) 'Technical Report IEEE 802.15.4 for Wireless Sensor Networks: A Technical Overview'.
- [8] Tooley, M. and Wyatt, D. (2009) *Aircraft Electrical and Electronic Systems Principles, operation and maintenance*. 1<sup>st</sup> Edition, USA: Elsevier Ltd.
- [9] Derriso, M.M. and Chang, F-K. (No Date) 'Future Roles of Structural Sensing for Aerospace Applications'.
- [10] Karl, H. and Willig, A. (2006) 'Protocols and Architectures for Wireless Sensor Networks'. John Wiley and sons Ltd.
- [11] G. A. Safdar and M. P. O'Neill (nee McLoone), "Performance Analysis of Novel Randomly Shifted Certification Authority Authentication Protocol for MANETs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 243956, 11 pages, 2009. doi:10.1155/2009/243956
- [12] Abbasi, A. A., Younis, M. (2007) 'A Survey on Clustering Algorithms for WSNs', *Computer Communications*, pp. 2826-2841, Science Direct, Published by Elsevier B.V.
- [13] Nakamura, E.F., Loureiro, A.A.F and Frery, A.C. (2007) 'Information Fusion for Wireless Sensor Networks: Methods, Models and Classifications', *ACM Computing Surveys*, 39(3), vol. 9, ACM Digital Library.
- [14] Hill, J.L. and Culler, D.E. (2002) 'Mica: A wireless platform for deeply embedded networks', *Micro IEEE*. 22 (6), pp. 12-24, IEEE Computer Society.
- [15] Hui, J.W. and Culler, D.E. (2008) 'Extending IP to Low-Power, Wireless Personal Area Networks', *Internet Computing, IEEE*, 12(4), pp 37-45, IEEE Computer Society.

# Modelling and experimental investigation of ferromagnetic material for angular defect detection

Dong Chang<sup>1</sup>, Xianzhang Zuo<sup>1</sup>, Yunze He<sup>2</sup>, Guiyun Tian<sup>2</sup>, Hong Zhang<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, Ordnance Engineering College  
Shijiazhuang, China

<sup>2</sup> School of Electrical, Electronic and Computer Engineering, University of Newcastle upon Tyne  
Newcastle upon Tyne, NE1 7RU, UK  
cdchangdong@126.com

**Abstract**—Ferromagnetic materials were widely used in various applications, on which crack is one of the most common defects caused by stress and environment factors, especially angular defects. In this paper, angular defects were modelled in the finite element simulation software COMSOL. Based on pulsed magnetic flux leakage testing, the distribution of magnetic flux leakage signal of defects were simulated. Compared with rectangular defects, some features were extracted to recognize the oblique crack. Then the experiments validated the correctness of simulation.

**Keywords**—ferromagnetic material; angular defect; Pulsed magnetic flux leakage; finite element simulation

## I. INTRODUCTION

As a strong magnetic material with good strength, hardness, ductility and toughness, ferromagnetic material has been widely used in the area of industrial productions such as petrochemical, aerospace transportation, machinery manufacturing and energy. Most of the key components of machinery are made of ferromagnetic materials. The crack may be easily generated by the effect of stress or other environments influence in these components. Those defects may cause unimaginable accident as hidden dangers in service. To safely and effectively make use of the materials, it requires to accurately grasp the performance of them in practice. Electromagnetic nondestructive testing is one of the most suitable methods for the detection of metal materials or components, which is based on the interaction between electromagnetic field and the conductive material. It includes of eddy current (EC), magnetic flux leakage (MFL), magnetic particle inspection (MPI), pulsed eddy current (PEC), and pulsed magnetic flux leakage (PMFL), etc.

As one of the most widely used electromagnetic nondestructive testing methods, PMFL is used in various applications; including inspection of storage tanks, pipelines and steel plates [1-2], but the most common use is in the inspection of oil and gas pipelines [3-4], where around 90% of pipeline inspection gauges (PIGs) use MFL technology [5]. MFL plays an important role in ferromagnetic materials for defect detection. At present, the crack detection of ferromagnetic materials has been focused on the rectangular crack. But the real cracks in applications have more complex geometry, such as the various angles of the oblique crack. I.Z. Abidin investigated eddy current distribution for angular defect

characterisation of conductive materials through simulation and experiment [6]. In this paper, the finite element simulation software COMSOL is used to simulate the distribution of angular defect's pulsed magnetic flux leakage signal, and it was validated through experiment.

## II PULSED MFL FOR ANGULAR DEFECT

PMFL is an electromagnetic nondestructive testing method which can detect surface and near surface defects of ferromagnetic materials. It developed from magnetic flux leakage and pulsed eddy current detection techniques. PMFL is effective to integrate the advantages of magnetic flux leakage technology and the Pulsed Eddy Current technology. Not only realized the magnetization of ferromagnetic materials to improve the penetration depth of excitation magnetic field, but also make good use of the square waveform and the rich frequency components which can provide information from different depths due to the skin effects.

The principle of PMFL to detect cracks was shown in Fig.1. The excitation signal is a certain duty cycle square wave, which is loaded to the excitation coil on the yoke to generate the transient magnetic field. If there is no crack in the material, the great majority of magnetic field lines cross the material, uniformly distributed in the material. If crack existed, the magnetic permeability of air in the crack was much smaller than the material itself. The reluctance increases at the crack. So that the magnetic field lines distorted along the crack and a part of them leak out the surface. Then the probe settled in the surface of test sample was employed to convert the magnetic flux leakage signal into the corresponding electrical signal. We can obtain the information of cracks by analyzing the electrical signal [7-8].

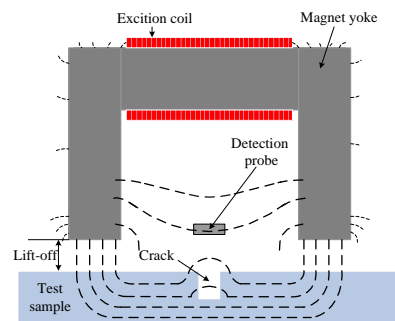


Figure 1. The principle of PMFL testing

### III SIMULATION

#### A. Simulation Setup

To analyze the influence of oblique crack on PMFL signal, four groups of angular model shown in Fig. 2 were established in COMSOL Multiphysics 3.5a. Fig. 2(a) is a common rectangular crack ( $90^\circ$ ), which was used to study the angular crack as a reference; Fig. 2(b) is a  $45^\circ$  angular crack; Fig. 2(c) is a group of cracks with the same width and various depth, the angles of  $22.5^\circ$ ,  $45^\circ$ ,  $67.5^\circ$ , which was used to analyze the influence of the angular defects' depth; Fig. 2(d) is a group of cracks with the same depth and various width, the angles of  $22.5^\circ$ ,  $45^\circ$ ,  $67.5^\circ$ , which was used to analyze the influence of the angular defects' width. All crack opening width are 0.5mm.

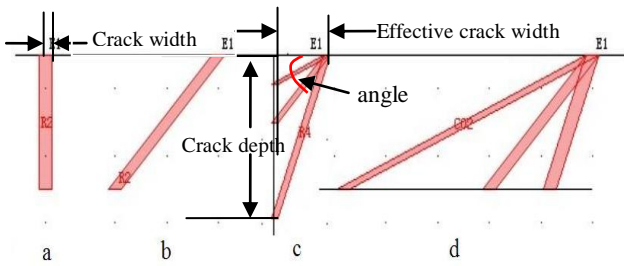


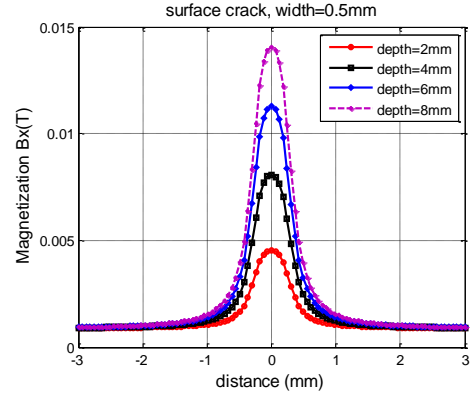
Figure 2. Angular defect models of ferromagnetic materials

#### B. Comparison angular defect with rectangular defect

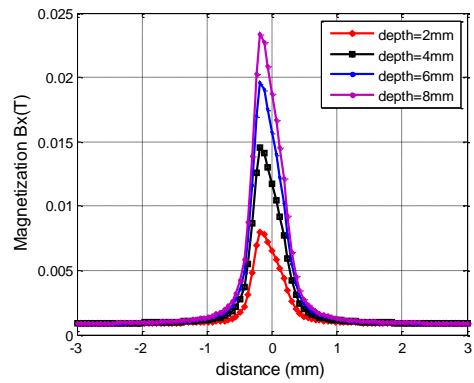
Rectangular and  $45^\circ$  oblique effective crack's depth changes from 2mm to 8mm, step with 2mm. The horizontal components of magnetic flux leakage of rectangular and  $45^\circ$  oblique crack were shown in Fig.3(a) and 3(b) respectively. The normal components were shown in Fig.4 (a) and 4(b).

Shown from the magnetic flux leakage waveform of rectangular defect, the horizontal component  $B_x$  is always positive and the peak position is the center of the crack. The normal component  $B_y$  has both positive and negative peaks, the peaks position are at the crack edge. The peak of the normal component is symmetric about the crack centerline and the value in the crack centerline is zero; the peak spacing (distance between peaks) and  $Vp-p$  (the difference between positive and negative peak) as a function of crack width and depth.

Because of their asymmetric structure, the oblique crack's waveform are asymmetric. Compared Fig.3(a) and 3(b), it can be found that the position of the peak of the horizontal component  $B_x$  shifted to the left of centerline. The offsets are the same as the same angle. The shift direction depends on the tilt direction of oblique crack, and its value increases with the crack depth increasing. Comparing Fig.4(a) and 4(b) shows the  $B_y$  of the angular crack has significant asymmetry on the positive peak and negative peak. The positive peak is prominently increased, on the contrary the negative peak reduced. The value of two peaks is also related to the tilt direction of oblique crack. The zero point is not in the centerline of the angular defect.

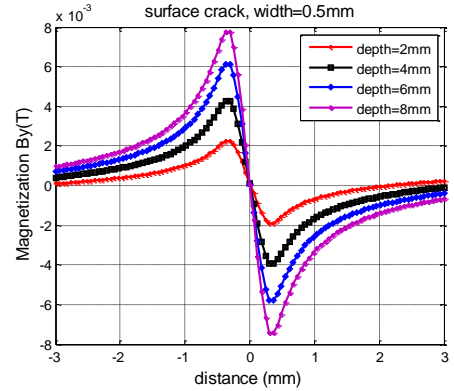


a) Rectangular

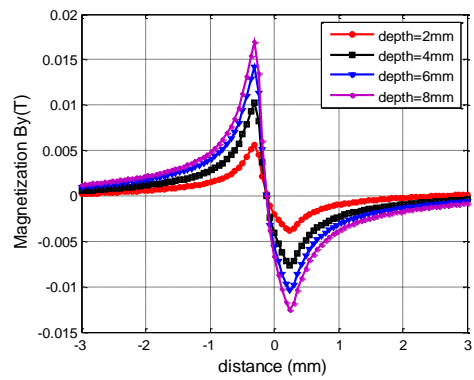


b)  $45^\circ$  crack

Figure.3 The  $B_x$  of rectangular and  $45^\circ$  crack



a) Rectangular



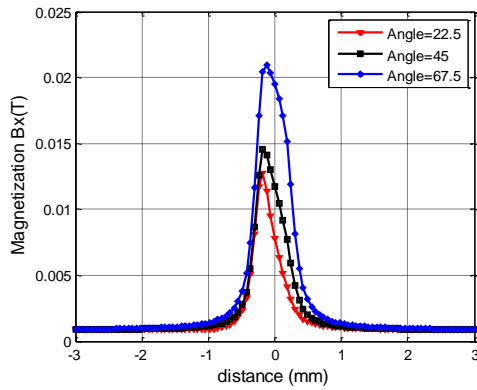
b)  $45^\circ$  crack

Figure.4 The  $B_y$  of rectangular and  $45^\circ$  crack

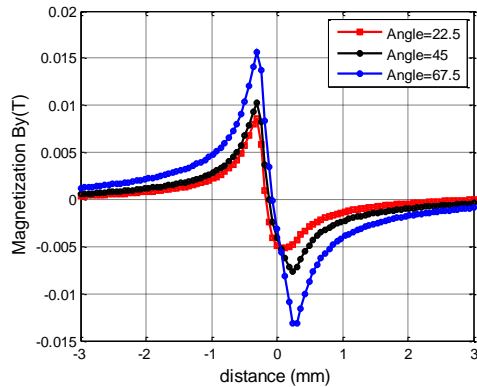
In summary, the difference of magnetic flux leakage signal between the oblique cracks with rectangular cracks were the shift of peak position in  $B_x$  and the asymmetry of positive and negative peak in  $B_y$ . We can qualitatively determine the tilt direction of the oblique crack by the position of peak occurrence and the value of positive and negative peak.

■ C. The identification of oblique crack's depth and width

The identification methods of angular crack's depth and width was researched with two different types of model, a group with the same effective width, depth changes and the other group with the same depth and effective width changes.

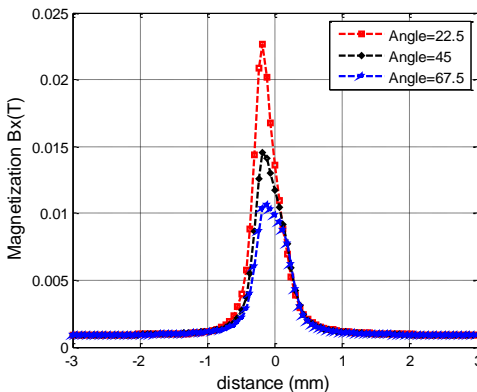


a) The  $B_x$  of the magnetic flux leakage

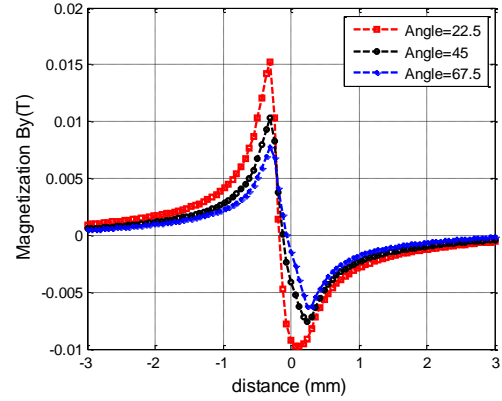


b) The  $B_y$  of the magnetic flux leakage

Figure.5 The angular crack with same effective width and different depth



a) The  $B_x$  of the magnetic flux leakage



a) The  $B_y$  of the magnetic flux leakage

Figure 6. The angular crack with same depth and different effective width

Studying for the model shown in Fig.2(c), the effective crack width is 4mm and the depth changes. Fig.5(a) and 5(b) show the pulsed magnetic flux leakage signals  $B_x$  and  $B_y$ . Considering to the model shown in Fig.2(d), the crack depth is 4mm with different angle. The pulsed magnetic flux leakage signals  $B_x$  and  $B_y$  were shown in Fig. 6(a) and 6(b).

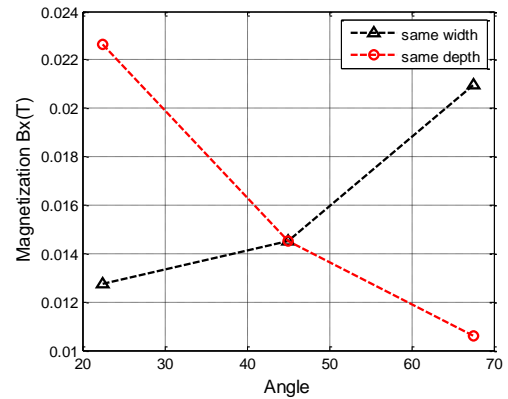
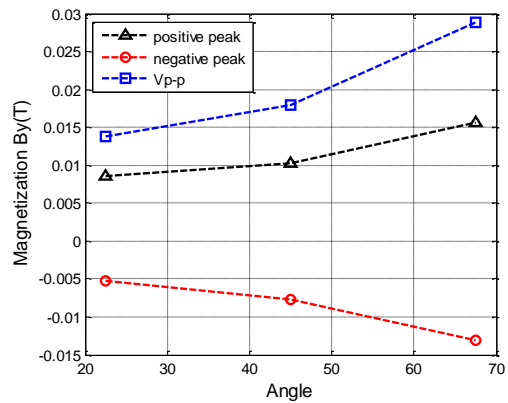
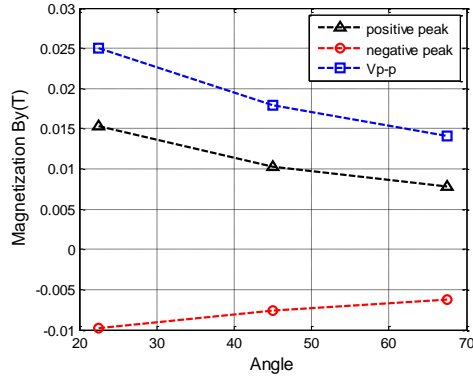


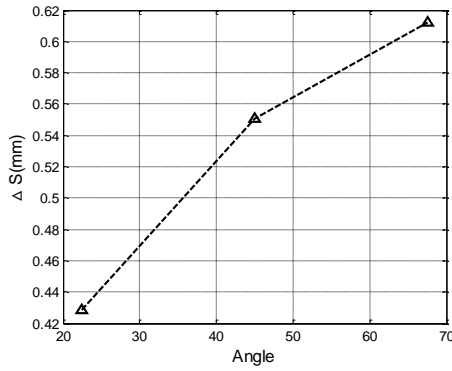
Figure 7. The  $B_x$  peaks



a) The same effective width , different depth



b) The same depth, different effective width



c) The peaks spacing of  $B_y$

Figure 8. The positive peak, negative peak,  $Vp-p$  and peaks spacing of  $B_y$

Fig.7 shows the curves of the  $B_x$  peak in two sets of model. It can be found that the peak increases with the depth and effective width of cracks increasing respectively. The value of peak is a function of effective crack width and depth. In Fig.8, the features of the normal component  $B_y$  were displayed respectively, the positive peak, negative peak and  $Vp-p$ . Comparing Fig.8(a) and 8(b), It can be pick up that it has the same trend when one of width and depth is constant, the other increases. So we can analyze the angular defect by measuring the normal component  $B_y$  of the magnetic flux leakage and extracting those features. Fig.8(c) shows that the peak spacing has a good linear relationship with the angle of cracks. Therefore, the angular can be recognized through detecting the peak spacing of  $B_y$ .

#### IV EXPERIMENT

##### A. Experimental Setup

In order to verify the correctness of the simulation results, the pulsed magnetic flux leakage detection system was established, shown in Fig.9. System consists of a square wave signal generator, amplifier, electromagnetic induction device, Hall detection devices, data acquisition card and PC. On the U- yoke, the coil is wrapped 400 turns on the two pole shoes. Hall detection probe was positioned in the middle of the two poles. The excitation square waveform current peak value is 600mA, the frequency is 50Hz, and the duty cycle is 0.5. The tests are carried out respectively on a steel sample with different artificial defects. Scanning direction is perpendicular to the defect. The peak value scanning curves are formed by extracting

the  $Vp-p$  of each scanning point's transient magnetic flux leakage signal.

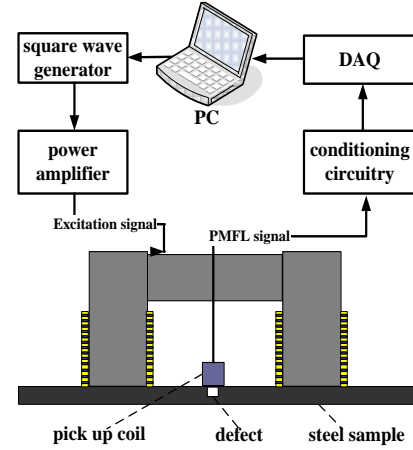
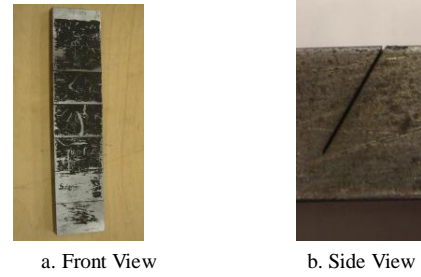


Figure 9. The schematic of the PMFL experimental system

##### B. Experimental results

On the steel sample, there are six cracks with the different angles; the dimension of steel is 350mm × 50mm × 15mm (length × width × depth ); the dimension of cracks are 50mm × 0.4mm ( length × width ), shown in Fig.10. The angle defects are shown in Tab.1. The scanning curves of different defects from the experiments are shown in Fig.11.

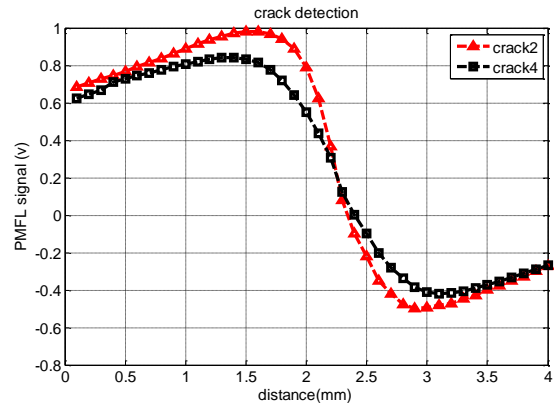


a. Front View b. Side View

Figure 10. Steel Sample with Angular Defects

Table 1. Angular defect Parameters

	Crack1	Crack2	Crack3	Crack4	Crack5	Crack6
Crack depth (mm)	4	5	10	5	1.5	0.5
Effective crack width (mm)	1.5	12	10	5	1.5	0.5
angle	67.5	22.5	45	45	45	90



a) Crack with same depth



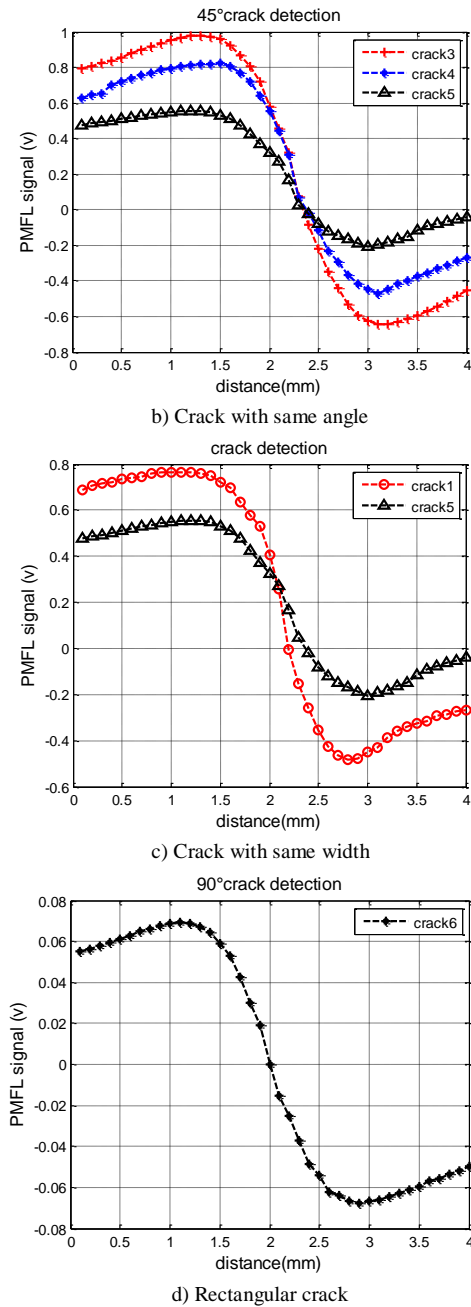


Figure 11. The PMFL signal of angular defects

Observing the shape of the curves, it can be found that the normal component of MFL signals is still bimodal curve. The asymmetry of positive and negative peak confirmed the presence of angular defect; according to which peak great recognize the tilt direction of the crack. Fig.11(a) shows the spacing of peaks is proportional to the width in the same crack depth and different width; Fig.11(b) shows the curve when the depth of cracks have the same angle changes; Fig.11(c) shows the peak-peak is a function of crack's depth when cracks have the same width and different depth; Fig.11(d) shows the curve of the rectangular crack that the signal distributed symmetrically. The experimental results matched with the simulations friendly.

## V. CONCLUSIONS

The simulation analysis and experimental validation

show that it is possible for detecting the angular crack by pulsed magnetic flux leakage. The angular crack can be quantitatively identified by the extracted features. To the normal component  $B_y$  of magnetic flux leakage signal, the asymmetry of the peaks can determine whether the crack is angular; the peak value is a function of crack depth; the position of zero point and which peak's value is great can recognize the tilt direction of crack; the spacing of peaks is a function of crack width and depth.

## REFERENCES

- [1] D. Jinfeng, K. Yihua and W. Xinjun. Tubing thread inspection by magnetic flux leakage, NDT & E International, Vol. 39, No. 1, 2006, pp. 53-56.
- [2] Y. Wang, Y. Xu, S. Ding, G. Dai, Y. Liu, Z. Yang and F. Liu. Numerical simulation and experiment on magnetic flux leakage inspection of cracks in steels, 17<sup>th</sup> World Conference on Nondestructive Testing, China, 2008.
- [3] Y. Zhang, Z. Ye and X. Xu. An Adaptive method for channel equalization in MFL inspection, NDT & E International, Vol. 40, No. 2, 2007, pp. 127-139.
- [4] X-B. Li, X. Li, L. Chen, P-F. Feng, H-D. Wang and Z-Y. Huang. Numerical simulation and experiments of magnetic flux leakage inspection in pipeline steel, Journal of Mechanical Science and Technology, Vol. 23, 2009, pp. 109-113.
- [5] A.A. Carvalho, J.M.A. Rebello, L.V.S. Sagrilo, C.S. Camerini and I.V.J. Miranda. MFL signals and artificial neural networks applied to detection and classification of pipe weld defects, NDT & E International, Vol. 39, No. 8, 2006, pp. 661-667.
- [6] I.Z. Abidin. Modelling and experimental investigation of eddy current distribution for angular defect characterisation[D], Newcastle University, 2010.
- [7] Tian Lu Chen, Gui Yun Tian, Ali Sophian, et al. Feature extraction and selection for defect classification of pulsed eddy current NDT [J]. NDT&E International, 2008, 41(6): 467-476.
- [8] Huang Zuoying, Que Peiwen, Chen Liang. 3D FEM analysis in magnetic flux leakage method[J]. NDT&E International, 2006, 39(1): 61-66.

# Defect depth effects in Pulsed Eddy Current thermography

Hong Zhang<sup>1</sup>, Guiyun Tian<sup>1</sup>, Yunze He<sup>1</sup>, Xianzhang Zuo<sup>2</sup>

<sup>1</sup>School of Electrical, Electronic and Computer Engineering, Merz Court, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, United Kingdom

<sup>2</sup>Department of Electric Engineering, Ordnance Engineering College  
Shijiazhuang, China  
hong.zhang1@newcastle.ac.uk

**Abstract**—In recent years, there has been an increasing interest in crack detecting and locating using Pulsed Eddy Current (PEC) thermography. PEC thermographic is an emerging integrative non-destructive testing (NDT) with the ability to inspect defects over large areas. The heating and diffusing results can be obtained easily and quickly from a thermal camera. In this paper, numerical modeling and experimental studies are applied to understand depth and tip effect on PEC thermography with different depths of defects, including transient heating propagation and magnetic flux distribution. This fundamental understanding of transient heating propagation and magnetic flux distribution will aid in the development of feature extraction and pattern recognition techniques for the quantitative analysis of PEC thermography images and crack characterisation.

**Keywords**—PEC; Thermography; Crack-tip; Finite element modelling; NDT; Magnetic flux

## I. INTRODUCTION

The use of thermography for NDT crack detection predates the development of the infrared camera by several decades. Early thermographic defect detection techniques date back to the 1960s, but it was the development of the infrared camera in the late 1970s which made it possible to directly detect the temperature contrast over large inspection areas [1, 2]. Further development of IR (Infrared) camera technology has led to increases in spatial resolution, frame rate and temperature sensitivity. Thermography NDT is a method, with the advantages of fast, non-contact, non-interaction, real-time measurement over a large detection area.

The major advantage of thermographic inspection over other detection techniques is that it able to inspect a large area (several m<sup>2</sup>) in a short time and is applicable to a wide range of materials, so the selection of the optimal excitation technique for the chosen application is important. Currently, thermography NDT has been widely used in aviation, aerospace, machinery, medical, petrochemical, power and other fields [3-6].

Although current heating excitation using flash lamps, etc. is still dominant, newer techniques such as eddy current and sonic excitation are gaining in popularity. In PEC thermography, a short burst (the inspection frequency typically is 50-500 kHz, the inspection period typically is 20ms-1s) of electromagnetic excitation is applied to the material under inspection, inducing eddy

currents to flow in the material. These eddy currents encounter a discontinuity, they are forced to divert, leading to areas of increased and decrease eddy current density. Areas where eddy current density is increased experience higher levels of Joule heating, thus these defects can be identified from the IR image sequence, include both during the heating and the cooling period[7].

Compare to the flash lamp heating [8-10], the PEC thermography has a direct interaction between the heating mechanism and the defects. This will result in a much greater change in heating around cracks, especially for vertical, surface breaking cracks. However, as with traditional eddy current inspection, the orientation of a particular defect with respect to induced currents has a strong impact, sensitivity decreases with cracks depth and the technique is only applicable to samples with a minimum level of conductivity.

The aim of this paper is to understand depth and tip effect on PEC thermography with different depths of cracks, including transient heating propagation and magnetic flux distribution. This paper has been divided into four parts. The first part gives a brief overview of the PEC thermography. The second part gives the numerical modeling and results by using COMSOL 3.4. The third part deals with the experiment system setup and results. Finally, the conclusions are given in the fourth part.

## II. NUMERICAL SIMULATION SETUP AND RESULTS

In this section, the simulation of magnetic field results will obtain by using COMSOL 3.4. The COMSOL is multi-disciplinary simulation software which permits designing and solving complex systems and structures by using Finite Elements Methods. The aim of these simulations is to analyse the comportment (form) of the magnetic flux distribution inside of the material (specimen) with different depths of defects.

### A. Simulation Setup

In order to obtain the magnetic flux distribution inside of the sample with defects. Magnetic flux is analysed in the simulation.

A U shape probe is used here to provide magnetic flux. The dimension of this U shape Yoke is shown in the table 1.

TABLE I. DIMENSIONS OF THE U SHAPE YOKE

	Height	Width	Length
<b>Probe</b>	200 mm	100 mm	400 mm
<b>Permeability</b>	$8.75 \times 10^{-4}$ H/m		
<b>Permittivity</b>	100 F/m		

The dimensions of metal sample are 40mm (length), 30mm (width) and 10mm (depth). Three different depths of cracks are created in this sample. Their depths are 2mm, 3.5mm and 7mm. The length and width of these cracks are the same 15mm (length) and 2mm (width). The current density in this metal is setup to 10 A/m<sup>2</sup>. This table below shows the dimensions of the specimen and the defects inside of the material.

TABLE II. DIMENSIONS OF THE SAMPLE AND DEFECTS

	Depth	Width	Length
Specimen	10 mm	150 mm	440 mm
First Defect	2 mm	15 mm	2 mm
Second Defect	3.5 mm	15 mm	2 mm
Third Defect	7mm	15 mm	2 mm

It can be seen from Figure 1, by using U shape yoke, the magnetic flux is forced vertically to the defect. The figure below shows the direction of the magnetic flux from the U shape probe to the metal plate specimen. There is no air gap between the material sample and probe.

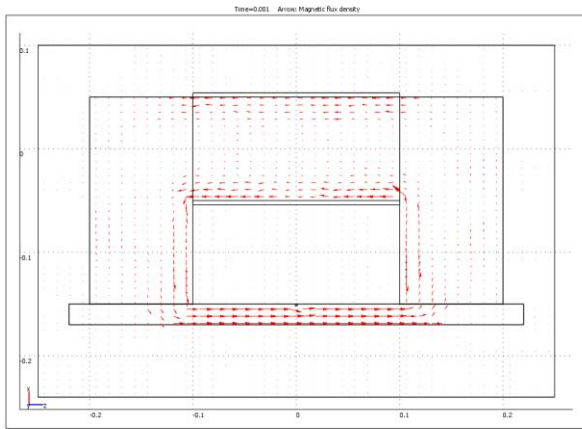


Figure 1. Simulated Magnetic flux distribution by using COMSOL

### B. Simulation Results

In this part of the paper, it is going to show the results from the simulations that were realized to determine the magnetic flux inside of the specimen and to obtain if the different depths of defects inside of the materials inspected can affect of the direction and magnitude of the magnetic flux on it.

It can be seen from the figures below, with the depth of the defect increasing, there will be more magnetic flux in both tips rather than in the bottom of the defect.

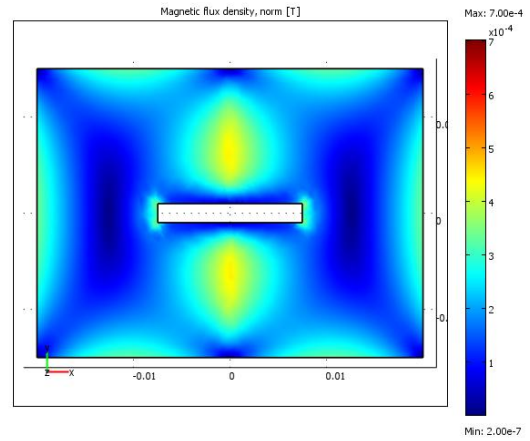


Figure 2. Top view of the defect with 2mm depth.

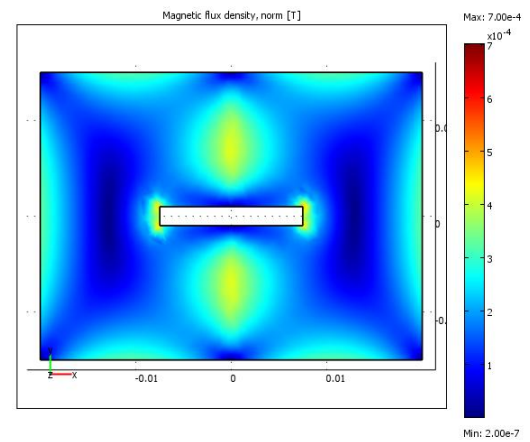


Figure 3. Top view of the defect with 3.5mm depth

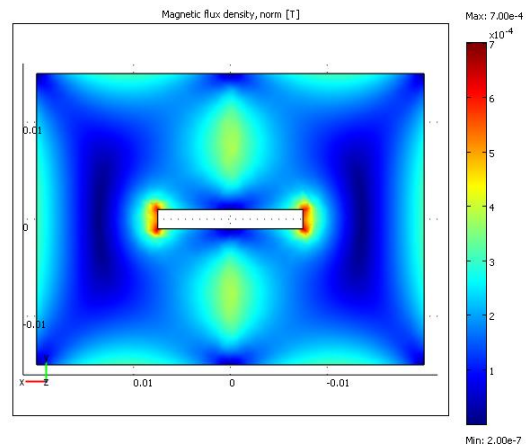


Figure 4. Top view of the defect with 7mm depth

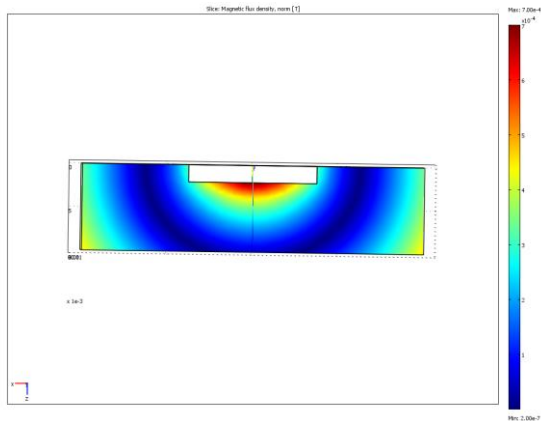


Figure 5. Cross section scheme of the defect with 2mm depth

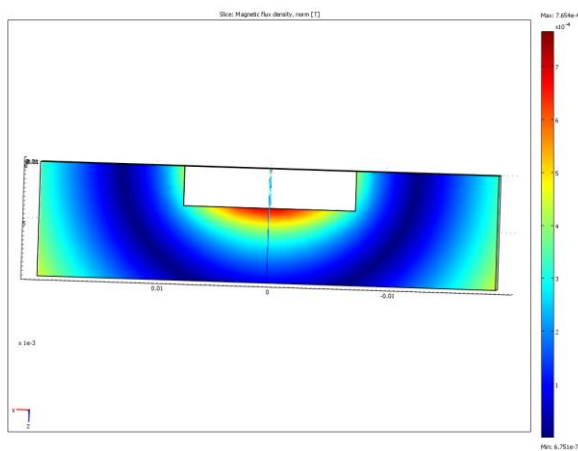


Figure 6. Cross section scheme of the defect with 3.5mm depth

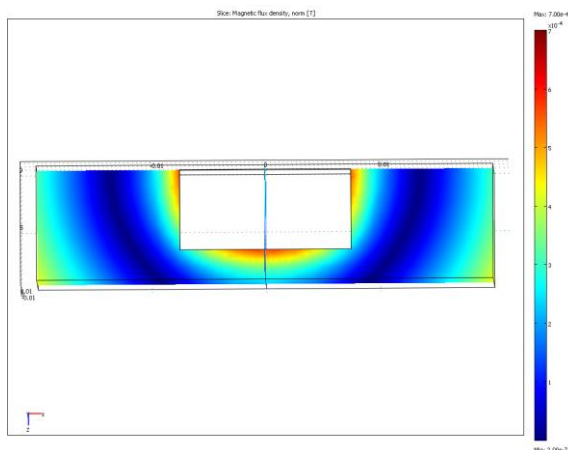


Figure 7. Cross section scheme of the defect with 7mm depth

It is notable by looking from the figures above, that the cross section shows that the presence of the magnetic field is increasing at the bottom of the defect inside of the material. The main difference between these simulation results is the depth of each defect. All defects are near to the surface of the material and there is no gap between the material and the excitation probe.

With these simulations, it shows that the magnetic reluctance (magnetic resistance) plays an important role in the defects position locating for the material detection. As result, by obtaining magnetic flux propagates inside of the material where the defects can be detected and located in material.

### III. EXPERIMENT SYSTEM SETUP AND RESULTS

The configuration of a pulsed eddy current thermography system, shown in Figure 8 is very simple and consists of an induction heating system which induces a copper coil is supplied with a eddy current of several hundred amps at a frequency of 50kHz – 1MHz from an induction heating system for a period of around 20ms – 1s. This induces eddy currents in the sample, which are diverted when they encounter a discontinuity leading to areas of increased or decreased heating. The resultant heating is measured using an IR camera and displayed on a PC.

For the experimental study, the coil here being an approximation of a Helmholtz coil (coil turns in image running roughly horizontally). The current induced in the surface of the test-piece will have been running approximately perpendicular parallel to the slot length. The current supplied to the coil was probably a full 380A, so the field strength at the centre of the coil would have been ~10-15mT.

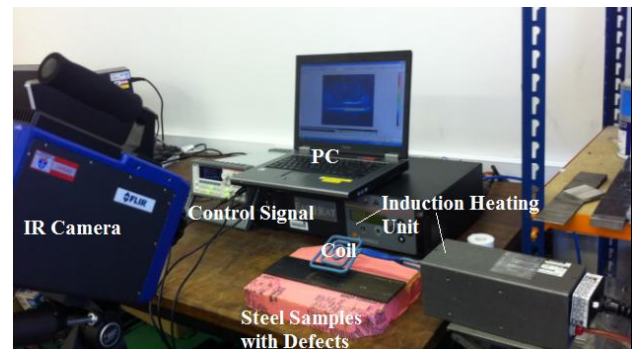


Figure 8. Basic configuration of a pulsed eddy current thermography system

There are 3 defects with different depths on the mild steel sample. This steel sample is 40mm\*30mm\*10mm. The length and width of these defects can be seen from table 1.

TABLE III. THE DEFECT PARAMETERS

Defect Number	Length (mm)	Width (mm)	Depth (mm)
1	15	2	7
2	10	2	3.5
3	10	2	2

The experiment results can be seen from Figure 9. It can be seen that only the tips of these defects are heated, not the whole side of these cracks. It is difficult to explain this eddy current thermal image using eddy current as the eddy current here is parallel to the defects. But this is understandable using magnetic flux distribution effect.

Therefore, numerical study in section 2 has been setup to prove this from magnetic field.

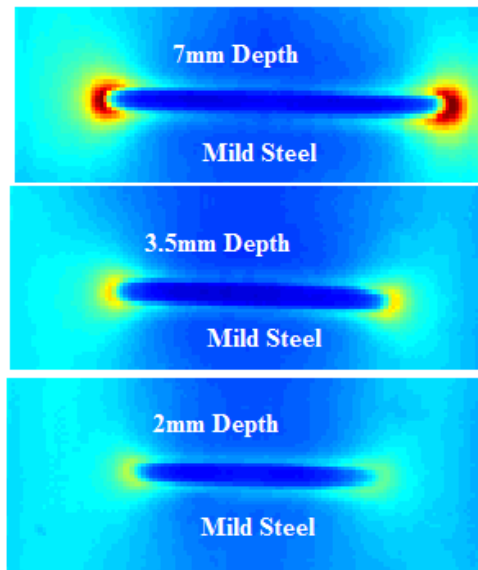


Figure 9. PEC thermography image of saw cuts depths

Compare these results with the experiment results, we can obtained there are good agreement between these results.

#### IV. CONCLUSIONS

The numerical simulation findings in this study provide a new understanding of PEC thermography with different depths of defects. The following conclusions can be drawn from the present study:

- For shallow surface cracks, when the magnetic field is perpendicular to the crack, the magnetic field lines mainly bypass the bottom of the defect; the magnetic field is still in accordance with the original distribution outside of defect.
- For deep surface crack, the magnetic field lines mainly bypass the crack's tips. Regardless of the depth of crack, because the magnetic resistance is much bigger in defect's air gap, the magnetic resistance is smaller in the material.

The findings of this study have a number of important implications for future thermograph imaging practice and defect locating. However, with a small sample size, caution must be applied, as the findings might not be

transferable to the other sample size. But this fundamental understanding of transient heating propagation and magnetic flux distribution will aid in the development of feature extraction and pattern recognition techniques for the quantitative analysis of PEC thermography images and defect characterisation.

#### REFERENCES

- [1] V. P. Vavilov, "Non-contact one-sided evaluation of hidden corrosion in metallic constructions by using transient infrared thermography," *Revista de Metalurgia (Madrid)*, pp. 235-242, 2003.
- [2] R. Arndt, B. Hillemeier, C. Maierhofer, et al., "Non-destructive detection of voids and inhomogeneities in building structures using impulse thermography," *Zerstörungsfreie Ortung von Fehlstellen und Inhomogenitäten in Bauteilen mit der Impuls-Thermografie*, vol. 81, pp. 786-793, 2004.
- [3] C. Garnier, M. L. Pastor, F. Eyma, et al., "The detection of aeronautical defects in situ on composite structures using non destructive testing," *Composite Structures*, vol. 93, pp. 1328-1336, 2011.
- [4] N. Tsopelas and N. J. Siakavellas, "Eddy current thermography in circular aluminium plates for the experimental verification of an electromagnetic-thermal method for NDT," *Nondestructive Testing and Evaluation*, vol. 25, pp. 317-332, 2010.
- [5] Y. Zhao, X. Guo, and M. Ren, "Lock-in thermography method for the NDT of composite grid stiffened structures," *Fuhe Cailiao Xuebao/Acta Materiae Compositae Sinica*, vol. 28, pp. 199-205, 2011.
- [6] D. G. Aggelis, E. Z. Kordatos, M. Strantza, et al., "NDT approach for characterization of subsurface cracks in concrete," *Construction and Building Materials*, vol. 25, pp. 3089-3097, 2011.
- [7] J. Wilson, G. Y. Tian, I. Z. Abidin, et al., "Modelling and evaluation of eddy current stimulated thermography," *Nondestructive Testing and Evaluation*, vol. 25, pp. 205-218, 2010.
- [8] A. Shima, R. Tsuchiya, T. Mine, et al., "Effects of thin film interference on junction activation during sub-millisecond annealing," *Japanese Journal of Applied Physics, Part 2: Letters*, vol. 46, pp. L427-L429, 2007.
- [9] A. Satta, A. D'Amore, E. Simoen, et al., "Formation of germanium shallow junction by flash annealing," *Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms*, vol. 257, pp. 157-160, 2007.
- [10] J. M. Laskar, S. Bagavathiappan, M. Sardar, et al., "Measurement of thermal diffusivity of solids using infrared thermography," *Materials Letters*, vol. 62, pp. 2740-2742, 2008.



# Parameters Influence in Steel Corrosion Evaluation Using PEC Thermography

Yunze He<sup>1</sup>, Guiyun Tian<sup>1</sup>, Liang Cheng<sup>1</sup>, Hong Zhang<sup>1</sup>, Paul Jackson<sup>2</sup>

<sup>1</sup>School of Electrical, Electronic and Computer Engineering, Newcastle University,  
Newcastle upon Tyne, NE1 7RU, United Kingdom

<sup>2</sup>International Paint Ltd. Gateshead,  
Tyne and Wear, NE10 0JY, United Kingdom  
y.he2@ncl.ac.uk

**Abstract**—Pulsed eddy current (PEC) thermography is proposed as a powerful Nondestructive Testing and Evaluation (NDT&E) technique, allowing operators to observe the heating developed from the eddy current distribution in a structure using infrared imaging, detecting defects over a relatively wide area within a short time. In this paper, PEC thermography is investigated to detect corrosion in structural steel components. The transient response at corruptions is a complex mix of many factors including electrical conductivity, permeability, thermal conductivity, heat capacity, density, depth, size, and exposure time which all have to be taken into account in the analysis of the PEC thermography. Firstly, the Finite Element Modeling (FEM) of corrosion with different parameters is established and numerical simulation is conducted in COMSOL Multiphysics. Next, structural steel (S275) samples with naturally produced corrosion are tested and experimental studies are carried out. The surface thermal profile and two related features are presented to evaluate the corrosion with different exposure time. The work shows that PEC thermography can be used for the detection and characterisation of corrosion through analysis of the surface thermal images and maximum temperature at heating and cooling stage.

**Keywords**—Pulsed eddy current thermography; Corrosion; Numerical simulation; Nondestructive testing and evaluation; Thermal profile

## I. INTRODUCTION

The major advantage of thermography over other NDT&E techniques is potential for rapid inspection of a large area within a short time [1]. To inspect defects over a large scale and at large stand-off distances, integration of thermography and other NDE approaches have been investigated, e.g. flash thermography [2], sonic thermography [3], laser thermography [4], optical thermography [5], and Pulsed Eddy Current (PEC) thermography [6, 7]. Pulsed Eddy Current (PEC) thermography, combining PEC and thermography, has its own advantages. The application of heat is not limited to the sample surface, such as in the flash thermography, rather it can reach a certain depth, which governed by the skin depth formula (Eq.1)... PEC thermography involves the application of high current EM pulse to the conductive material under inspection for a short period (typically less than 1s). When the eddy currents encounter a discontinuity (e.g. crack, delamination, etc), they are

forced to divert, leading to regions of increased and decreased eddy current density. Then, increased eddy current density will lead to higher levels of Joule (Ohmic) heating. Thus, the defect can be identified from a characteristic heat distribution in the thermal image/video. After a period of eddy current heating, the defect also affects the heat diffusion in the cooling phase. Therefore the mixed phenomena of induction heating dominating the heating phase and heat diffusion dominating the cooling phase and their specific behaviour is used for the quantitative NDE (QNDE) of defects.

Steel components are used in many industry sectors, such as shipping, off shore oil and gas production, power plants, and coastal industrial plants. Because of the hostile environment, corrosion frequently occurs on steel components. More seriously, corrosion will form holes and wall thinning under a coating, which will cause operational safety problems. Depending on the corrosion environment there are two kinds of corrosion: corrosion in natural environments (atmospheric, soils, freshwater, seawater, and microbial) and in industrial environments. According to the mechanism of corrosion, there are also two kinds of corrosion: chemical corrosion and electrochemical corrosion. According to the characteristics of the corrosion breakdown, corrosion can be categorised into: uniform corrosion (general corrosion), galvanic corrosion, pitting corrosion, crevice corrosion, filiform corrosion, intergranular corrosion, stress corrosion cracking (SCC), and corrosion fatigue. In this paper, PEC thermography technique is investigated for atmospheric exposed electrochemical general corrosion characterisation in steel material. Corrosion in steel causes chemical and physical changes, which change electrical conductivity, permeability [8], thermal conductivity, heat capacity, depth, and density. All these aspects have to be taken into account in PEC thermography.

The rest of the paper is organised as follows: Firstly, a model of steel with corrosion is established and numerical simulation is conducted in the Section II. Next, PEC thermography set-up is introduced and experimental study is carried out and the surface thermal profile and two related features are presented to evaluate the corrosion in Section III. Finally, conclusions and future work are outlined in Section IV.

## II. NUMERICAL STUDIES

### A. FEM model

As shown in Fig. 1, the 3D Finite Element Modelling (FEM) is established in COMSOL Multiphysics, which consists of steel, corrosion and coil. According to coordinate system, the steel size is constant as 150×60×3mm and 30mm long (in y axis) corrosion with varied width and depth are investigated. Table I shows the material parameters used in the model [8, 9]. It is noticed that one parameter is various and other parameters are constant when the influence of this parameter is studied.

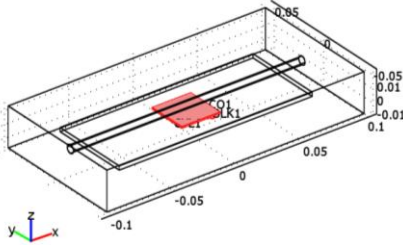


Figure 1. The 3D finite element modelling

TABLE I. PARAMETERS OF STEEL AND CORROSION

Parameters	steel	corrosion
Conductivity, (S/m)	4.68e6	0.75e6
Relative Permeability	60	4
Density, (kg/m <sup>3</sup> )	7850	5242
Heat capacity, (J/kg·K)	475	100
Thermal conductivity, (W/m·k)	44.5	0.6

The excitation frequency and current are set as 256 kHz and 380A to match the experimental set-up. The heating time is 200ms and the cooling time is 300ms. Fig. 2 and Fig. 3 show the thermal images at the end of heating and cooling phase respectively. It can be noticed that the temperature rise at corrosion is larger than that at steel, which is contributed by all parameters discussed previously. Next, we will find out the influence of each parameter.

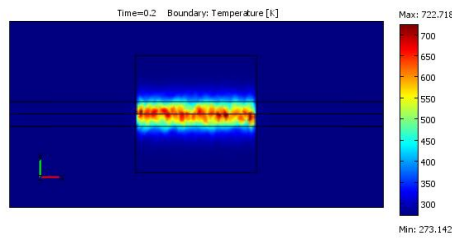


Figure 2. The thermal image at the end of heating (0.2s)

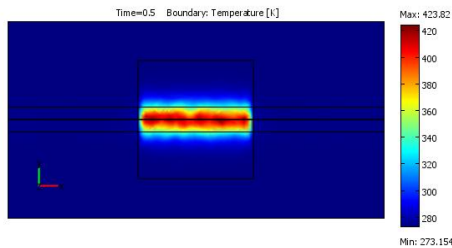


Figure 3. The thermal image at the end of cooling (0.5s)

### B. Influence of electrical conductivity and permeability

Firstly, according to Skin Effect, the penetration depth  $\delta$  of eddy current in the conductive material can be calculated by

$$\delta = 1/\sqrt{f\sigma\mu} \quad (1)$$

where  $f$  is excitation frequency,  $\sigma$  is the electrical conductivity, and  $\mu$  is the permeability of the material under inspection.

Secondly, Joule heating is caused by resistive heating from the eddy currents. The generated resistive heat  $Q$  is proportional to the square of the eddy current density  $J_s$  or electric field intensity  $E$ . The relationship between  $Q$ ,  $J_s$  and  $E$  is governed by following equation.

$$Q = \frac{1}{\sigma} |J_s|^2 = \frac{1}{\sigma} |\sigma E|^2 \quad (2)$$

Therefore, electrical conductivity and permeability will affect the temperature rise on the surface of the corrosion.

In the simulation for investigating the conductivity influence, the other parameters are constant (permeability is 60 and other parameters are shown in Table I). In the simulation of different permeability, the other parameters are constant (conductivity is 4.68e6 and other parameters are shown in Table I). The corrosion size is 30×30×1 mm.

Fig. 4 shows the maximum temperature at the corrosion at 0.2s against different electrical conductivities and permeabilities. With the conductivity decreasing from 100% to 6.67% of the original value 4.68e6 S/m, max T increases from 516.843 K to 1786.677 K. On the contrary, with the permeability decreasing from 100% to 6.67% of the original value 60, max T has a little decrease from 516.843 K to 474.077 K.

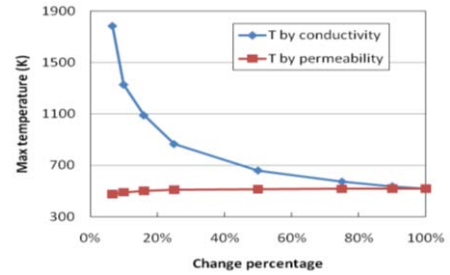


Figure 4. The maximum temperature at the end of heating

Fig. 5 shows the maximum temperature at the corrosion at 0.5s against different conductivities and permeabilities. With the conductivity decreases from 100% to 10% of 4.68e6, max T increase from 337.094 K to 590.693 K. On the contrary, with the permeability decreases from 100% to 10% of 60, max T decrease from 337.094 K to 332.373 K, which is a weak change.

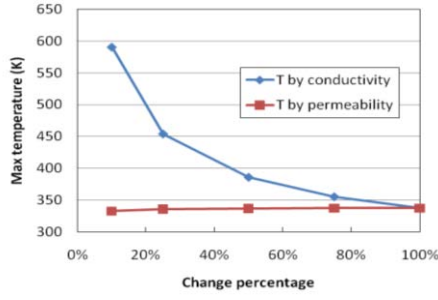


Figure 5. Maximum temperature at the end of cooling

### C. Influence of thermal conductivity, heat capacity, and density

The heat conduction equation of a specimen caused by a Joule heating source  $Q$  is governed by

$$\rho C_p \frac{\partial T}{\partial t} - \nabla(k \nabla T) = Q \quad (3)$$

where  $\rho$ ,  $C_p$ ,  $k$  are density, heat capacity and thermal conductivity respectively.

Fig. 6 shows the maximum temperature at the corrosion at 0.2s and 0.5s against different thermal conductivity. With the thermal conductivity decreases from 44.5 to 0.6, max T at 0.2s increase from 303.916 K to 721.59 K, max T at 0.5s increase from 282.36 K to 423.82 K, and the temperature difference between 0.2s and 0.5s has a rise.

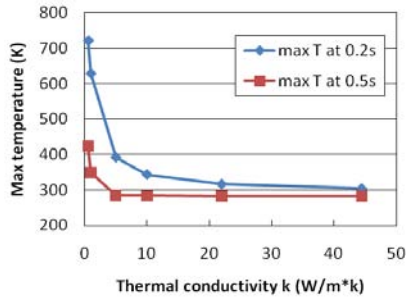


Figure 6. The maximum temperature at the end of heating and cooling against different thermal conductivity

Fig. 7 shows the maximum temperature on the corrosion at 0.2s and 0.5s against different heat capacity. With the heat capacity decreases from 475 to 100, max T at 0.2s increase from 427.617 K to 721.59 K, max T at 0.5s increase from 365.707 K to 423.82 K, and the temperature difference between 0.2s and 0.5s has a rise.

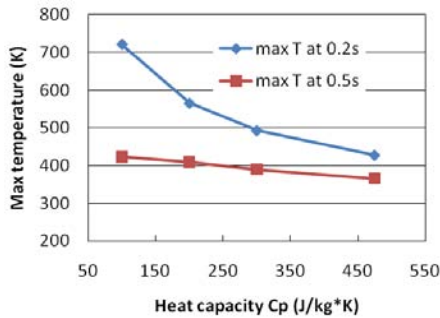


Figure 7. The maximum temperature at the end of heating and cooling against different heat capacity

Fig. 8 shows the maximum temperature on the corrosion at 0.2s and 0.5s against different density. With the density decreases from 7850 to 4000, max T at 0.2s increase from 625.668 K to 795.647 K and the temperature difference between 0.2s and 0.5s has a rise. On the contrary, max T at 0.5s has a little change.

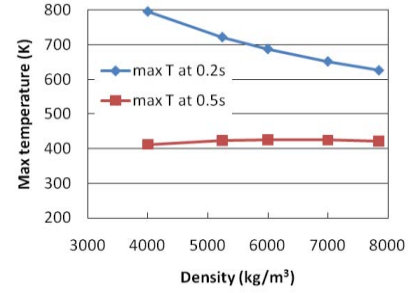


Figure 8. The maximum temperature at the end of heating and cooling against different density

### D. Influence of lift-off

Fig.9 shows the transient temperature change curves on the corrosion with different lift-off. With the lift-off increases, max T at 0.2s and 0.5s, the raising slope in heating stage, and falling slope in cooling stage have a decrease. That means the sensitivity decreases.

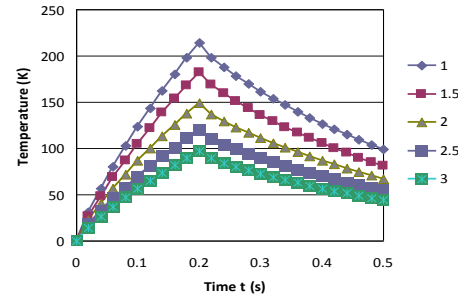


Figure 9. The temperature change of corrosion with different lift-off

The normalisation using division of the transient temperature by the temperature at maximum heating, is usually used in PEC thermography to illuminate lift-off influence [11]. The normalised curves of temperature change in Fig. 9 are shown in Fig. 10. Clearly, the overlap of normalised curves infers that the normalisation eliminates the lift-off effect.

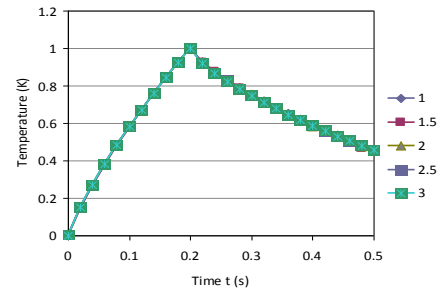


Figure 10. The normalised temperature change of corrosion with different lift-off

### E. Influence of corrosion depth and size

Fig. 11 shows the temperature change of corrosion with different depth from 0.3mm to 0.8mm. With the depth increase from 0.3mm to 0.8mm, max T at 0.2s and 0.5s has a monotonic rise.

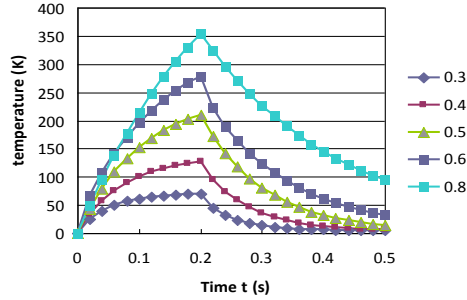


Figure 11. The temperature change of corrosion with different depth

The normalised curves of temperature change in Fig 11 are shown in Fig. 12. Obviously, with the depth increase from 0.3mm to 0.8mm, the time of arriving at the same normalised temperature increases in the heating stage and the time to the same normalised temperature increases in the cooling stage. That is because the corrosion depth means the distance of thermal diffusion.

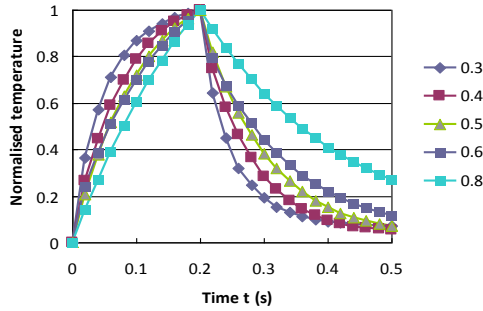


Figure 12. The normalised temperature change of corrosion with different depth

Fig. 13 shows the maximum temperature on the corrosion at 0.2s and 0.5s against different corrosion widths. With the corrosion width increases from 10mm to 50mm, max T at 0.2s and at 0.5s have a small change. In that, the PEC thermography response is not sensitive to corrosion width.

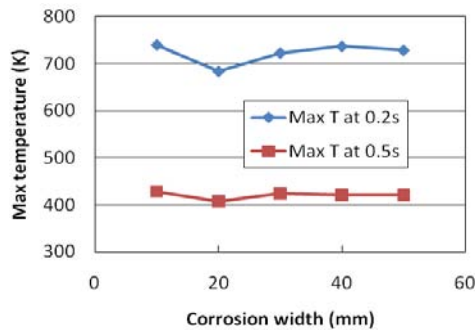


Figure 13. The maximum temperature at the end of heating against different thermal conductivity

## III. EXPERIMENTAL STUDIES

### A. Set-up

The experimental setup is illustrated in [10] and shown in Fig. 14. An Easyheat 224 from Cheltenham Induction Heating is used for coil excitation. The rectangular coil is constructed from 6.35 mm high-conductivity hollow copper tube. The Easyheat has a maximum excitation power of 2.4 kW, a maximum current of 400 Arms and an excitation frequency range of 150 kHz - 400 kHz (380Arms and 256 kHz are used in the experiments). In general, high excitation frequencies will lead to high thermal contrast (or high temperature rise). The time domain information will allow the derivation of defect profile information. The system has a quoted rise time (from the start of the heating period to full power) of 5 ms, which was verified experimentally. Water cooling of coil is implemented to counteract direct heating of the coil.

The SC7500 is a Stirling cooled camera with a 320 x 256 array of 1.5 - 5  $\mu$ m InSb detectors, shown in Figure 14. The camera has a sensitivity of <20 mK and a maximum full frame rate of 383 Hz. The maximum 383 Hz frame rate provides 1 frame every 2.6 ms, with the option to increase the frame rate with windowing of the image. In our studies, a high-speed thermal camera is used for investigation including feature optimisation for QNDE. However, in real applications, it may be unnecessary to use such a high-end camera.



Figure 14. The photo of PEC thermography set-up

### B. Corrosion sample

Mild steel (S275) samples with corrosion are provided by International Paint®. Fig. 15 shows a photograph of a typically corroded steel sample. As shown in Table II, sample 18, 14 and 4 provide corrosion with different exposure times from 1 month to 6 months.



Figure 15. The photo of sample 18

TABLE II. SPECIFIC DETAILS OF STEEL SAMPLE

No.	Condition	Defect shape	Exposure time
-----	-----------	--------------	---------------



18	Uncoated	Square 3cm*3cm	1 month
14	Uncoated	Square 3cm*3cm	3 months
4	Uncoated	Square 3cm*3cm	6 months
5	Uncoated	Square 3cm*3cm	10 months

### C. Results and discussion

Fig. 16 shows the thermal images of sample 18 and 14 at the end of heating. Fig. 17 shows the thermal images of sample 4 at the end of heating and cooling. Firstly, max temperature of corrosion at the end of heating and cooling is much higher than that of steel, which is accordant with the simulation results in Fig. 2 and Fig. 3. Secondly, temperature of corrosion is very light in some area and not very light in some area, which is obvious in thermal images of sample 4 in left figure of Fig.17. The reason of this phenomenon is that there are different products and surface morphology of corrosion. The different products have the different conductivity, permeability, density, conductivity, and heat capacity, which all can affect the surface PEC thermography response as well as the corrosion depth. Therefore, the surface thermal profile is presented to evaluate the corrosion.

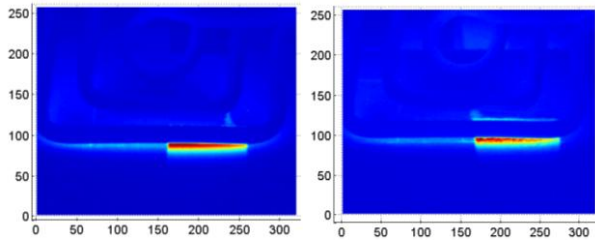


Figure 16. The thermal images of sample 18 and 14 at 0.2s

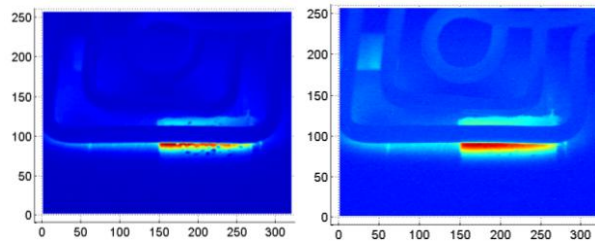


Figure 17. The thermal images of sample 4 at 0.2s and 0.5s

The thermal image of part of corrosion on sample 14 at the end of heating in is show in Fig. 18(a). The line 1 is on the surface of corrosion. Fig. 18(b) shows the surface thermal profile of line 1, which is parallel to coil. Because of the different types of products in corrosion and the surface roughness, the surface thermal profile of corrosion can be used to evaluate the corrosion. The features peak-peak (PP) and standard deviation (Std-dev) of profile line are presented. The PP and Std-dev of surface thermal profile in Fig. 18(b) is 435 DL and 70.52 DL, respectively.

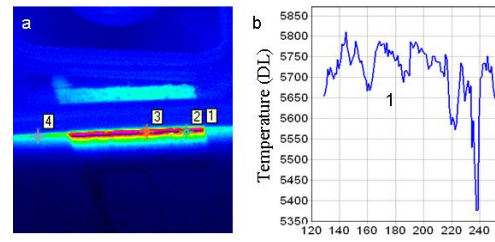


Figure 18. (a) the thermal images of corrosion at the end of heating phrase; (b) the surface thermal profile.

The dependence of presented features to exposure time of uncoated corrosion is shown in Fig. 19. Clearly, the two features have an increase with the extension of exposure time. The power function is widely used in long-term corrosion ( $t > 1$  year) development [12]. In previous work, the early stage corrosion ( $t < 1$  year) development has been characterised using power function [13] as shown in equation (4)

$$F = At^n \quad (4)$$

Where,  $t$  is the exposure time in months,  $F$  is feature after  $t$ ,  $A$  is that in the first month, and  $n$  is constant, which is less than 1. Corrosion rate can be derived from the first derivative of equation (5)

$$v = nAt^{n-1} \quad (5)$$

In the development of corrosion, the rate of supply of oxygen to the corroding surface will decline as the corrosion product layer builds up over time [14]. Then, the corrosion rate will decrease as exposure time  $t$  increases, which means the slope of the fitted lines decreases over exposure time. However, it is noticed that the slope of PP has a sudden increase from 6 months to 10 months. The reason is that some rust loose and flake off on the 10 months exposed corrosion, which can increase the surface roughness. Fig. 20 shows corrosion with 3 and 10 months in exposure time. As shown in Fig. 20(a), there is almost no rust flake off. But in Fig. 20(b), some rust loose and flake off as it expands.

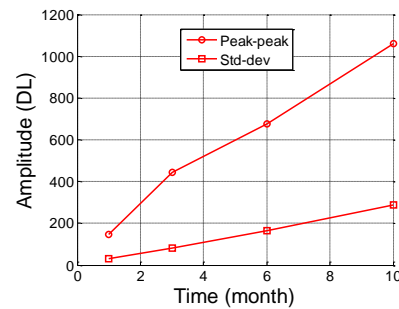


Figure 19. the dependence of features to exposure time

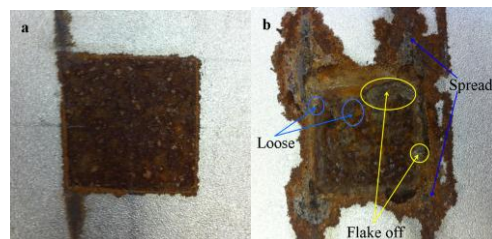




Figure 20. the photos of corrosion with exposure time of (a) 3 months, (b) 10 months.

#### IV. CONCLUSIONS

In PEC thermography, the response of corrosion is a complex mix of many factors including electrical conductivity, permeability, thermal conductivity, heat capacity, density, and corrosion morphology, which all have to be taken into account in the analysis of the PEC thermography. Firstly, the model of corrosion with different parameters is established and numerical analysis is studied in COMSOL Multiphysics. Next, structural steel (S275) samples with naturally produced corrosion are tested and experimental studies are carried out:

- The electrical conductivity, permeability, thermal conductivity, heat capacity, density of corrosion is smaller than that of steel. Through numerical simulations, it is known that decrease on electrical conductivity, thermal conductivity, heat capacity, and density will result in increase on temperature on the surface of corrosion. Decrease on permeability will lead to a small decrease in temperature. Numerical and experimental studies show the temperature of corrosion is higher than that of steel at the heating and cooling stage.
- Corrosion depth has a big influence on the temperature change. Maximum temperature has a monotonic increase with depth increase from 0.3mm to 0.8mm. On the contrary, corrosion width has a small influence. Therefore, the surface roughness or morphology of corrosion should be taken into account in experiment studies. With the exposure time increase, the roughness or morphology of corrosion has a more and more remarkable influence on the thermal images.
- The surface thermal profile and two related features are presented to evaluate the corrosion with different exposure time. The two features have an increase with the extension of exposure time.

The future work includes that parameters influence on transient temperature change curve and corrosion characterisation under coatings.

#### ACKNOWLEDGMENT

The authors would like to thank the International Paint® for experimental samples. Mr. Yunze He would like to thank China Scholarship Council for sponsoring his visit study to University of Newcastle and Professor Xianzhang Zuo from Ordnance Engineering College, China for useful discussion.

#### REFERENCES

- [1] N. P. Avdelidis, B. C. Hawtin and D. P. Almond, "Transient thermography in the assessment of defects of aircraft composites", *NDT&E International*, vol. 36, pp. 433-439, 2003.
- [2] R. Mulaveesala and S. Tuli, "Theory of frequency modulated thermal wave imaging for non-destructive sub-surface defect detection", *Applied Physics Letters*, vol. 89, No.19, pp. 191913 - 191913-3, 2006.
- [3] M. Morbidini and P. Cawley, "The detectability of cracks using sonic IR", *Journal of Applied Physics*, vol. 105, issue 9, pp. 093530- 093530-9, 2009.
- [4] S. E. Burrows, A. Rashed, D. P. Almond and S. Dixon, "Combined laser spot imaging thermography and ultrasonic measurements for crack detection", *Nondestructive Testing and Evaluation*, vol. 22, issue 2-3, pp. 217-227, Jun. 2007.
- [5] C. M. Zöcke, Quantitative analysis of defects in composite material by means of optical lockin thermography, Dr. Ing. Dissertation, Saarbrucker Reihe Materialwissenschaft Und Werkstofftechnik, Dec 2009.
- [6] I. Z. Abidin, G. Y. Tian, J. Wilson, S. Yang and D. Almond, "Quantitative evaluation of angular defects by pulsed eddy current thermography", *NDT & E International*, vol. 43, issue 7, pp. 537-546, Oct. 2010.
- [7] S. Yang, G. Y. Tian, I. Z. Abidin and J. Wilson, 'Simulation of edge cracks using pulsed eddy current stimulated thermography', *Journal of Dynamic Systems, Measurement, and Control*, vol. 133, 2011.
- [8] Y. Gotoh, H. Hirano, M. Nakano, K. Fujiwara, and N. Takahashi, "Electromagnetic nondestructive testing of rust region in steel," *IEEE Transaction on Magnetics*, 2005, vol. 41, pp.3616-3618.
- [9] S. B. N. Laaidi, A. Elbaloutti, "Thermal and thermographical modeling of the rust effect in oil conduits," *ECNDT*, 2010.
- [10] J. Wilson, G. Y. Tian, I. Z. Abidin, S. Yang and D. Almond. "Pulsed eddy current thermography system development and evaluatio", *Insight*, vol. 52, pp 87-90, 2010.
- [11] L. Cheng and G. Tian, "Surface Crack Detection for Carbon Fibre Reinforced Plastic (CFRP) Materials Using Pulsed Eddy Current Thermography," *Sensors Journal*, IEEE, vol. PP, pp. 1-1, 2011.
- [12] W. Hou and C. Liang, "Atmospheric corrosion prediction of steels," *Corrosion*, vol. 60, pp. 313-322, 2004.
- [13] Y. He, G. Tian, H. Zhang, and P. Jackson, "Atmospheric Corrosion Characterisation using Pulsed Eddy Current Feature Optimisation," *IEEE Sensors Journal*, vol. submitted, 2011.
- [14] R. E. Melchers, "Transition from marine immersion to coastal atmospheric corrosion for structural steels," *Corrosion*, vol. 63, pp. 500-514, 2007.

# The Effect of Metallic Substance on the Read Range of UHF Passive RFID System

Chencho<sup>1,2</sup>, Dr. Justin Champion<sup>2</sup>, and Prof. Hongnian Yu<sup>2</sup>

1. College of Science and Technology, Royal University of Bhutan, Phuentsholing, Post .Box #450, Bhutan.

2. Faculty of Computing, Engineering and Technology, Staffordshire University, Beacon, Beaconside, Stafford, Staffordshire, ST18 0AD, United Kingdom.

Email: tshencho\_753@live.com, j.j.champion@staffs.ac.uk, h.yu@staffs.ac.uk

**Abstract**—Read Range is one of the important performance metric of passive RFID system. It depends on the parameters of both the reader and a tag. It also depends on the application environment. Passive RFID system is often used where a tag is tagged to a metallic objects or are used where there are metallic surfaces in the proximity. This paper gives how the metallic surface affects the read range of UHF Passive RFID system. This was carried out to see how the metallic plate in the proximity effect the reader and tag antenna parameters and how these parameters affect the read range. The effect on antenna parameters were studied under two cases. Firstly, using a metallic plate at different distances from the antenna and secondly using plates of different sizes. The simulation of antenna was carried out in the NEC2D and the effect on read range is carried out in MATLAB. The work was carried out during my term of 20 months at Staffordshire University funded by e-link.

**Keywords**- Antenna, Metallic, Patch, Passive, Read Range.

## I. INTRODUCTION

This paper is divided into four sections. Section I gives a brief introduction to UHF passive RFID system, Section II gives the effects of metallic surface on the gain and impedance of the antenna. Section III shows how antennas gain; antenna impedance affects the read range of passive UHF RFID system. Section IV end with the conclusion. RFID systems are used to identify, locate and tracking purposes using radio frequency. It consists of three components; reader, tag and an external database as shown in figure 1.

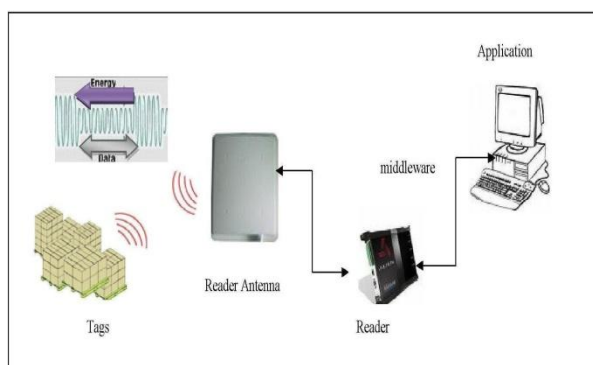


Figure 1: RFID Network.

The tag contains the identification number and unique

code. It consists of an antenna which is connected to the chip. The reader reads the data from the tags which are in its read range. The middleware is the part of system software and retrieves the data from the reader and sends it to the external database. There are two types of RFID system; Active and Passive RFID system. In the passive RFID system, the tag utilizes the time varying radio frequency from the reader as the power source to energize itself and to communicate with reader. The tag does not contain any onboard power source unlike active RFID system.

A UHF passive RFID system operates between the frequencies ranges of 860MHz - 960MHz. The operating frequencies are different in different countries. It depends on the radio regulations of their own country. The RFID based on UHF and higher frequency use far field communication [1]. In the radar system technology, the reflected electromagnetic wave depends on the object's reflection cross section. Objects are in resonance with the wave that hits them have a particularly large reflection cross section [2]. This is same case for the RFID system operating in far field region. For electrically small antennas ( $D \ll \lambda$ , where  $D$  is the maximum dimension of antenna and  $\lambda$  is the wavelength), far field region is commonly given as [1]:

$$r = \frac{\lambda}{2\pi} \text{ m.} \quad (1)$$

where ' $r$ ' is the distance from antenna  
For  $D > \lambda$ ,

$$r = \frac{2D^2}{\lambda} \text{ m} \quad (2)$$

The far field is limited by the amount of energy and sensitivity of the reader receiver. The reflected energy is of small value and decays with the factor of  $1/r^4$  [3].

### System handshake in UHF passive RFID system

Typical system handshake in UHF passive RFID system is as follows [4]:

1. The reader continuously emits time varying carrier RF sine wave. Modulation in the field will occur if there are tags in its read range.

2. Upon receiving the threshold power by the tag, it begins clocking down its data to an output transistor which is normally connected across the antenna input.
3. The output transistor sequentially shunts the coil corresponding to the data clocked out of the memory array. This shunting causes dampening of the carrier wave resulting in a change in amplitude of the carrier.
4. The reader detects the modulated data and process the resulting bit stream according to the encoding and data modulation methods used.
5. The middleware retrieves and filters the data feed to the application software.

The above process described in sending data from tag to reader is called backscattering.

The performance of RFID is characterized by its read range. The read range of passive RFID system depends on the many parameters like reader and tag properties and also on the environment depending on the nature of surrounding. Section II shows how the gain of the antenna and antenna impedance are affected by metallic patch of different sizes and when a metal patch is placed at a different distant from the reader.

## II. EFFECT OF METALLIC SUBSTANCE IN THE PROXIMITY ON THE READ RANGE

Metallic substances are good conductor of electricity and heat. They are the electropositive elements (donates electrons to form positive ions) which can be melted or hammered into thin sheets or drawn into wires. Radio frequency wave signals are attenuated when there is an obstacle in the propagation path. The attenuation depends on the properties of substances like metals and nonmetals. Metals have very high degree of attenuation. They reflect the incident waves from its surface with the phase reversal [5]. Therefore the presence of metallic substances blocks the signals from the source and the affect depends on the conductivity of the metallic surfaces or the substances. The metal plates or surfaces with the higher conductivity blocks the signal more than the ones with less conductivity.

The performance of Passive RFID system is strongly dependent on the material of tagged object made from different substances like metal, plastics, wood, glass, Meat etc. Studies have found that the performance of the RFID system degrades when passive tags are tagged to the objects made from metallic substances and containers containing liquid. Also the read range decreases when there are metallic substances in the proximity of the RFID system. The effects of nearby object on the read range of RFID system using several types of tags was carried out in [6] and found that there is decrease in electric field near the surface to meet boundary condition requirement;

this was the main reason for the decrease in read range. Similar a study was carried out on the proximity effects of metallic environments on the high frequency RFID reader antenna in [7]. They have designed reader antenna and studied the performance in the various configuration of metallic plate in terms of resonant frequency, field intensity and field distribution. They found out that the size, the orientation and the distance of the metallic plate from the reader antenna greatly affects the read range of whole RFID system by shifting up the resonant frequency.

The EM waves near the metallic surface are explained in [8]. It has shown that the performances of any RFID tag that depends on either the normal component of the magnetic field or tangential component of electric field will degrade when attached to or have metallic substances in the proximity.

In the study [7], it was found that the resonant frequency of the antenna in the proximity of the metal plate is always shifted with a corresponding change in the field intensity. The effect differs depending on the position of the metal plates. In their study they found that the effect of back-placement plate is more severe than the bottom. This is because the antenna properties such as the input impedance, directivity, radiation pattern and efficiency gets changed; the change in tag antenna impedance will then lead to the shift in resonant frequency of the tag and secondly it causes impedance mismatch [8]. The frequency shift of the matching circuit can be calculated using the equation

$$fr = \frac{1}{2\pi\sqrt{LC}} \text{ Hz} \quad (3)$$

Where  $fr$  is the resonant frequency,  $L$  and  $C$  are the inductance and capacitance of the circuit respectively. It can be seen from the equation that the resonant frequency depends on the change in impedance of the antenna resulting in the detuning and read range degrades.

For the maximum power transfer between the tag antenna and the chip, the impedances should match i.e. the chip impedance is complex conjugate of the tag antenna impedance. When the tag is attached to or is placed near the metallic surface, the tag antenna impedance is affected and this causes mismatch in impedance. This will also affect the bandwidth over which good performance is obtained [8].

A simple half wave dipole antenna is designed in NEC2D taking frequency 890MHz. Figure 2 shows the normalized gain of the designed antenna. Following examples shows how the gain of the half wavelength dipole antenna gets reduced and changes the impedance value in presence of metallic surfaces in the proximity.

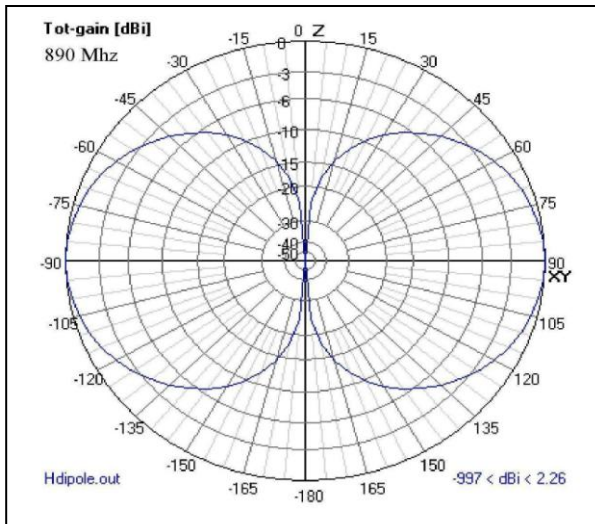


Figure 2: Normalized Gain of the Simple half dipole Antenna

The study on the read range in presence of metallic patch is carried taking two cases.

*Case 1:* When a metallic plate is placed at different distances from the antenna.

Figure 3 shows how the gain or radiation pattern gets changed when a metallic surface is placed at different distances from the antenna; the figure shows that the gain decreases as the separation distance between the antenna and metallic surface (170 mm X 20 mm) decreases. The metal plate is placed at a distant of 30mm, 25mm, 20mm and 10mm from the antenna.

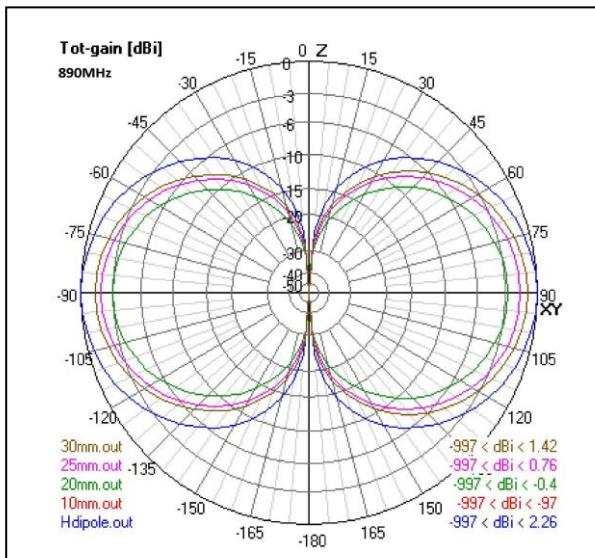


Figure.3: Radiation when the metal patch is placed at different distant from the antenna.

*Case 2:* When metal plate of different size metal plate are used.

Metal patch of different sizes, 200 x 20mm, 160 x 20mm,

120 x 20mm and 80 x 20mm are used in the simulation. Figure 4 shows the effect of metal patch size to the antenna gain and it's seen that the gain decreases with the increase in the metal patch size.

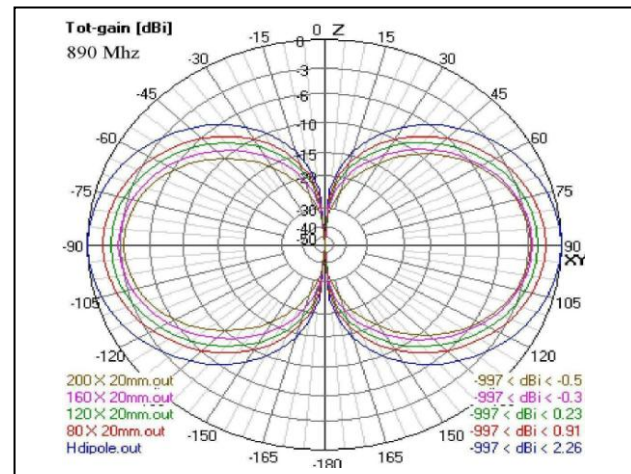


Figure 4: Radiation when the metal patches of different sizes are placed near the antenna.

The reflection and SWR also changes when there is metal patch near the antenna. This shows that there is change in impedance of the antenna in presence of metallic substances as shown in Figure 6. It can be seen that there is increase in reflection coefficient and SWR when a metallic surface of (200 X 20) mm is placed at the distance of 20 mm from the antenna.

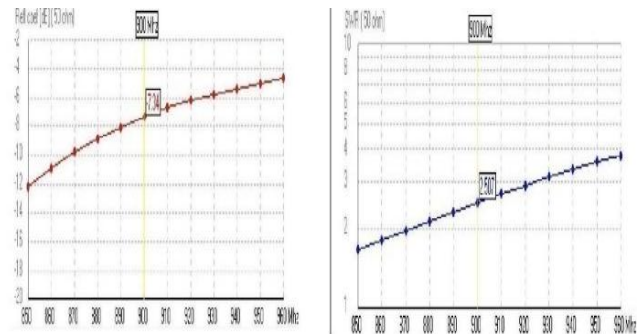


Figure 5: Half Wave Dipole reflection co-efficient and SWR without metal surface.

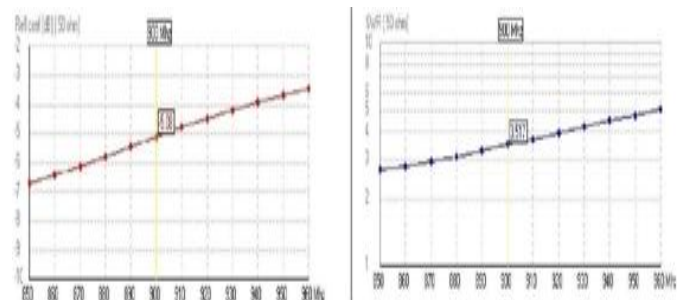


Figure 6: Half Wavelength dipole reflection co-efficient and SWR with Metal Surface.



The impedance for the above designed antenna is changed upon placing metallic patch at a distant of 20mm from the antenna as can be seen in Figure 6.

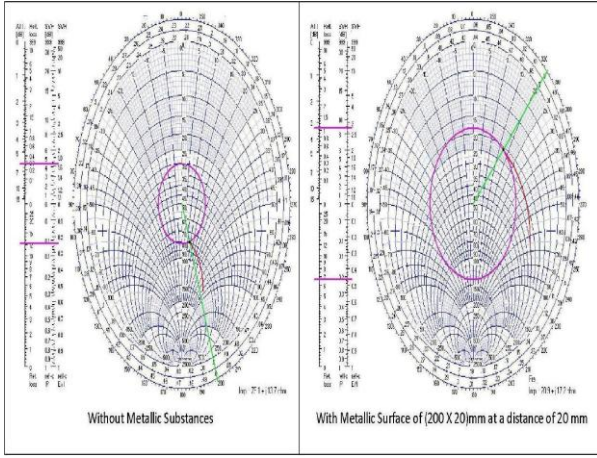


Figure 7: Change in impedance of antenna due to metallic patch of 200 x 20 mm at a distance of 20mm

It can be concluded that both reader and tag gain is affected by metallic substances and the affect depends on the size and the distance of the metallic substances from the antenna. Also the presence of metallic patch changes the impedance value of the antenna. The changes in these three parameters does affect the read range and section III shows how these parameters affects the read range of Passive UHF RFID tags.

### III. EFFECT OF ANTENNA GAINS AND IMPEDANCE MISMATCH ON THE READ RANGE.

Both the reader and tag antenna will have effects on the read range of RFID system. Figure 7 explains how the read range increases with the increase in the tag antenna gain with different transmitting power of reader. The transmitting power of reader depends on the country's radio regulators. It differs from nation to nation. It shows that the read range can be increased if they tag antenna is of larger gain.

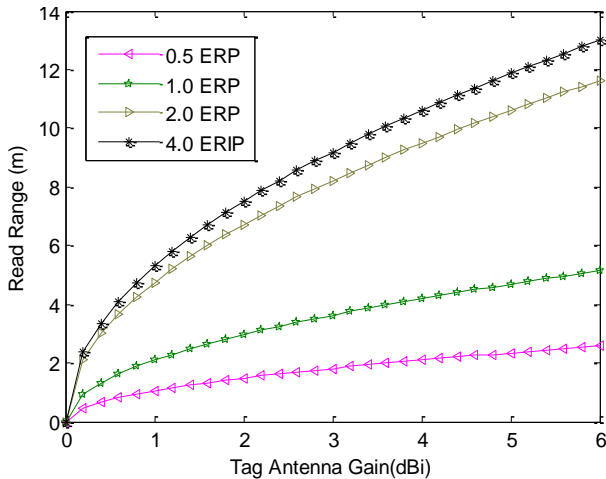


Figure 8: Effect of Tag antenna Gain on the Read Range

The read range is calculated taking -10dBm threshold power of tag and operating frequency at 890MHz.

Similarly Figure 8 shows how the read range of the passive UHF RFID system varies with the increasing gain of the reader obtained at different values of reader sensitivity.

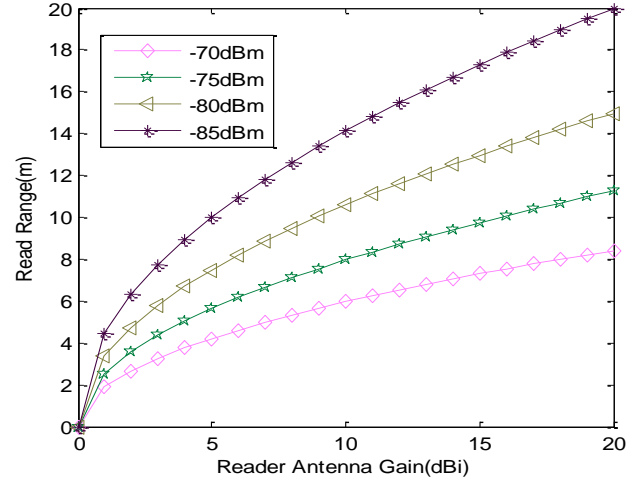


Figure 9: Effect of Reader Antenna gain on the Read range.

#### Impedance Matching

The impedance matching between the tag antenna and the chip is characterized by the power transmission coefficient. The power transferred from the tag antenna to the chip is maximum when the impedance matching is perfect i.e. when perfect complex conjugate impedance match between tag antenna and chip takes place. Consider Tag antenna impedance of  $Z_A = R_A + jX_A$  and Chip Impedance be  $Z_C = R_C + jX_C$ .

Power transferred from tag antenna to chip is maximum when  $Z_A = Z_C^*$

(4)

The amount of power transferred to the chip is given [9] by

$$P_{chip} = P_{antenna} \times \tau \quad (5)$$

where,  $\tau$  is the power transmission coefficient and is given by

$$\tau = \frac{4R_A R_C}{|Z_C + Z_A|^2} \quad (6)$$

$\tau$  value is '1' when the impedances are matched perfectly. Therefore read range with different values polarization loss factor and power transmission coefficient is given by [9]:



$$r = \frac{\lambda}{4\pi} * \sqrt{\frac{EIRP * G_r * p\tau}{P_{tag}}} m \quad (7)$$

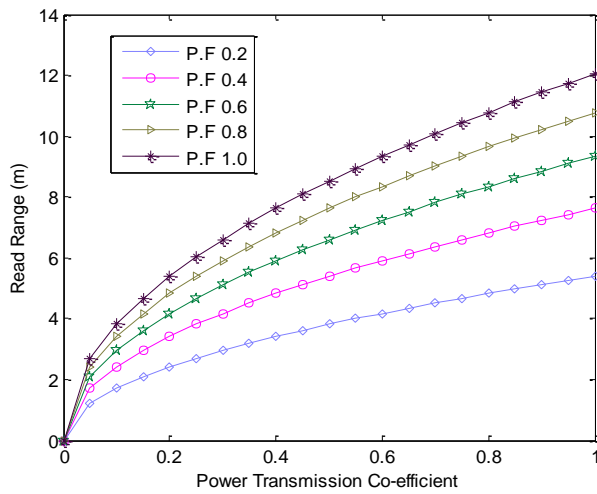


Figure 10: Effect of Impedance on the Read Range

#### IV. CONCLUSION

It can be concluded from the results above that the read range is greatly affected by the presence of metallic substance in the RFID passive system. The metals affects the antenna gains of both the reader and also changes the value of impedance and this in turn affects the overall read range. Therefore the antenna designer should take these factors into consideration and design an antenna suitable to operate when used in different environment.

#### ACKNOWLEDGEMENT

This work is being carried out during the term in Staffordshire University, UK funded by the e-link; an EU project aims to develop new and enhance existing co-operative teaching and research links and to establish a sustainable virtual learning platform to continue collaborative activities.

#### REFERENCES

- [1] Lehpamer, H., 2008, *RFID Design Principles*, USA: Artech House Inc.
- [2] Finkenzeller, K. 2003. *RFID Handbook- Fundamentals and applications in contactless smart cards and Identification*, 2<sup>nd</sup> Ed. New York: Wiley and Sons LTD.
- [3] Want, R., 2006. *An introduction to RFID technology*. IEEE Pervasive computing, 5(1), pp. 25-33.
- [4] Sorrells, P., 1999. *AN-680 Passive RFID basics*. Microchip Technology Inc., [Online]
- [5] Ukkonen K., 2005, *Effect of metallic plate size on the performance of micro strip patch type tag antennas for passive RFID*. Antenna and wireless Propagation letter. IEEE[online]
- [6] Daniel M. D.; Weigand, S.M., 2005. *Environmental Effects on RFID Tag Antennas*. In: Microwave Symposium Digest, IEEE MTT-S International.
- [7] Qing X. & Chen Z. N., 2007. *Proximity Effects of Metallic Environments on High Frequency RFID Antenna: Study and Application*. Antennas and Propagation, 55(11), pp. 3105-3111. [Online] 5 June.
- [8] Ahson S. & Ilyas M., 2008. *RFID Handbook, Applications, Technology, Security and Privacy*. Boca Raton: CRC Press, Taylor & Francis Group, 6000 Broken Sound Packway NW, Suit 300.
- [9] Eunni, M.B, 2004. *A novel planar micro strip antenna design for UHF-RFID*. Master thesis.

# Electromagnetic NDT and Condition Monitoring – A Personal View

Xiandong Ma  
Engineering Department  
Lancaster University  
Lancaster, UK LA1 4YR  
Email: xiandong.ma@lancaster.ac.uk

**Abstract**—This paper will review the research and development of eddy current NDT testing, condition monitoring and the associated instruments, which the author has been closely involved over the past decade. Eddy current testing and imaging is firstly introduced, focusing on property measurement of porous metals and imaging liquid metals. Condition monitoring is then presented from fundamental aspects in the high voltage insulation diagnosis to applications in large-scale generators and power plants. Finally, the challenges we will face in these areas are summarised.

**Keywords**—Nondestructive testing (NDT); eddy current; electromagnetic induction imaging (EMI); condition monitoring; power plant

## I. INTRODUCTION

Nondestructive testing and evaluation is an interdisciplinary field of study which is concerned with the development of analysis techniques and measurement technologies for the quantitative characterization of materials, tissues and structures by noninvasive means. Ultrasonic, radiographic, thermographic, electromagnetic, and optic methods are employed to probe interior microstructure and characterize subsurface features. The traditional areas are for instance flaw detection, material characterization and structural health monitoring.

Historically, condition monitoring and analysis was performed by highly experienced maintenance individuals who applied lessons learned through years of practice. Over the past decade, there has been much interest shown internationally in the development and deployment of new technologies and intelligent systems for use in the energy generation fields. A growing selection of techniques have been researched and developed for monitoring and diagnosing the key components and equipment of a power plant during their operations. These techniques embrace electrical, mechanical, thermal, electromagnetic, acoustic, optical, chemical and meteorological phenomena to yield monitoring data for collection and desirable instruments for processing and interpreting the gathered data.

This paper will review the research and development of eddy current NDT testing, condition monitoring and the associated instruments, which the author has been closely involved over the past decade. In Section II, eddy current testing and imaging is introduced, focusing on property measurement of porous metals and imaging liquid metals. In Section III, condition monitoring is presented from

fundamental aspects in the high voltage insulation diagnosis to applications in large-scale generators and power plants. The challenges we will face in these areas will be summarised in Section IV.

## II. EDDY CURRENT TESTING AND IMAGING

For a fundamental eddy current testing, two inductive coils are used with one acting as an exciter and the other a detector. The excitation coil with a number of turns is excited by an AC source, which generates a changing magnetic field in its vicinity. This time-varying magnetic field interacts with the test sample and then induces eddy currents. The eddy currents in turn generate a secondary magnetic field. The interaction between two fields alters the distribution of the magnetic flux, resulting in an apparent change of the coil voltage. By measuring this coil voltage change, the properties such as the electrical conductivity and magnetic permeability of the material can be measured by using analytical and/or experimental methods.

With more inductive coils, it is possible to tomographically image the distribution of materials inside a region of interest. The sensor array can contain excitation and detection coils which can either be dedicated to a particular function (exciter or detector) or operate in both modes. By energising an excitation coil with an AC signal, a measurement can be obtained from detection coils. This kind of projection is continued until last excitation sensor is excited and measurement is taken. The measured data is then manipulated using mathematical inversion techniques to create an image of the internal object distribution.

### A. Forward Problems

Eddy current problems are related to the forward and inverse problems. The forward problem is to predict the sensor outputs for a given material distribution with known electromagnetic properties. This problem can be formulated in terms of the magnetic vector potential  $\mathbf{A}$  for the sinusoidal waveform excitation cases using complex phasor notation

$$\nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{A} \right) + i\omega\sigma\mathbf{A} = \mathbf{J}_s \quad (1)$$

where  $\sigma$  is electrical conductivity,  $\mu$  is magnetic permeability,  $\omega$  is the applied frequency and  $\mathbf{J}_s$  denotes the source current passing through the excitation coil. Having obtained the vector potential  $\mathbf{A}$ , the electric field

intensity  $\mathbf{E}$  can be calculated through  $\mathbf{E} = -\partial\mathbf{A}/\partial t$ . The induced voltage in detection coil is thus computed by taking the line integral of the vector  $\mathbf{E}$  around the coil loop.

The direct solutions to (1) are not straightforward. Generally, analytical solutions are suitable for ideal geometries and require simplified assumptions for the geometry. The numerical approaches, such as finite-element methods (FEM), are used to calculate approximately the forward problems in the general case. There are some commercial FEM packages available to solve the eddy current forward problems, such as COMSOL and ANSYS.

In the imaging context, the sensitivity is defined as the change in the induced voltage on the pairs of coils with respect to a change in the conductivity of the conducting region. It can be calculated using the  $\mathbf{AA}$  formulation below

$$\frac{\partial V_{ij}}{\partial \sigma_k} = -\omega^2 \frac{\int \Omega_k A_i \cdot A_j}{I_i I_j} \quad (2)$$

where  $V_{ij}$  is the induced voltage of coil pair  $i$  and  $j$ ,  $\Omega_k$  is the volume of element number  $k$  in the conducting region  $\sigma_k$ , and  $I_k$  and  $I_j$  are excitation currents for the coils. Sensitivity matrixes are widely used to solve the inverse problems in image reconstruction, as they describe the unique conductivity distribution to pixel perturbations for a given sensor array.

### B. Inverse Problems

The inverse problem involves converting the measured data into properties of material for fundamental NDT testing or back into an image of the original material distribution for more sophisticated NDT imaging. It is worth emphasising that the reconstruction problem of conductivity mapping is an ill-posed and ill-conditioned problem as the number of pixels in an image is usually much larger than the number of the limited measurements. This problem is further complicated by the soft field effect, whereby the object material affects both the magnitude and direction of the interrogating field. Therefore, the measured induced voltage is a nonlinear function of electrical conductivity of the test objects. In general for NDT imaging the object space is divided in a number of small elements for which linear equations can be assumed. Consequently, the system can then be represented by a large number of linear equations, which can be treated using matrix manipulation techniques.

### C. NDT of Porous Metals

The application of eddy current inspection techniques to porous metals is relatively new. The induced eddy currents flowing in the foam samples are affected significantly by the foam properties and consequentially measurement of the impedance change on the detection coil(s) permits the metal foams to be characterised.

Normally, different shapes of sensor coils are constructed purposely to accommodate different shaped

samples. For example, pancake-type surface coils are used to inspect plate, sheet or irregularly shaped samples. Encircling coils are used primarily for inspecting rods, tubes, cylinders, or wire in manufacturing applications. When the area to be tested is large, pancake-type surface coils are chosen to reduce the testing time whereas coils as small as practical are required to detect small cracks. In practical measurements, double-coil arrangements are preferred, where one coil is used for excitation purpose with a separate secondary coil used for detection. The separation of excitation coil and detection coil can avoid measurement problems due to overheat of the excitation coil.

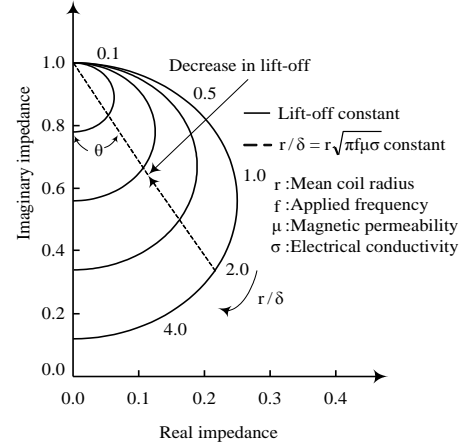


Figure 1 Normalised impedance curves for planar coil varying with reference numbers  $r/\delta$  and lift-off

The normalised impedance analysis has been widely used for the analysis of eddy current signals in a complex-plane diagram. It is defined as the ratio of the measurement coil impedance due to the presence of the test object and the coil impedance as measured in the air. Fig. 1 illustrates the variations of real and imaginary impedance components for electrically conductive (nonmagnetic) materials under different values of reference number and lift-off. The reference number is defined as  $r/\delta$ , i.e., the ratio of mean coil radius  $r$  and skin depth  $\delta$ . As can be seen from this illustration, impedance magnitude varies with the lift-off (coil-to-sample spacing) significantly whilst the phase data is independent with the test geometry. Therefore, phase data is very useful in analysing eddy current signal as it is only related with the material under test and less affected by the test geometries [1].

The electrical conductivity  $\sigma$  can be evaluated using the recalculated values of  $X(\theta)$  by

$$\sigma = \frac{1}{\omega \mu (rX(\theta))^2} \quad (3)$$

where  $X(\theta)$  can be calibrated based on the impedance curves of bulk materials as their electrical conductivity is already known. Fig. 2 shows variations of electrical conductivity of the aluminium foams with both porosity and pore size. In summary, higher porosity leads to lower equivalent electrical conductivity due to the decreasing of the metal volume ratio in the foams. The pore size

determines the amount of air trapped, the average wall thickness between pores and the degree of interconnectivity between pores, thus affecting the equivalent conductivity of the foams as well [2]. The technique can also be used for measurement of magnetic permeability of porous Fe samples [3].

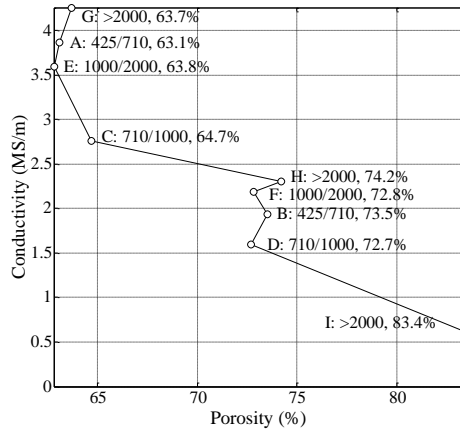


Figure 2 Electrical conductivity varying with both porosity and pore size of the aluminium foams

#### D. Liquid Metal Imaging

The mathematical concept of tomography was first proposed early in the 19th century. The basic aim of modern tomography is to determine the cross-sectional distribution of materials of interest by a set of measurements using sensors that are distributed around the periphery of a process. Electrical tomography has opened up many applications. This is mainly because this sensing technique is non-intrusive and non-invasive; has a relatively high imaging speed, low cost and can be used in the harsh operating conditions, e.g. in the cases where temperature and pressure are extremely high [4].

Fig. 3 shows a tomographic imaging system, consisting of three main subsystems, namely a sensor array, a conditioning electronics unit and a PC equipped with a DAQ board. The image reconstruction algorithms based on the SIRT (simultaneous iterative reconstruction technique), Tikhonov regularization and SVD (singular value decomposition) have been selected for this system [5]. Fig. 4 shows several image results of aluminium samples with a relatively complex shape. Although these linear reconstruction techniques generally produce qualitative images of the property distribution, the image contrast produced is indeed capable of showing the correct object position by the brightest pixel and the relative size of the objects by pixel intensity

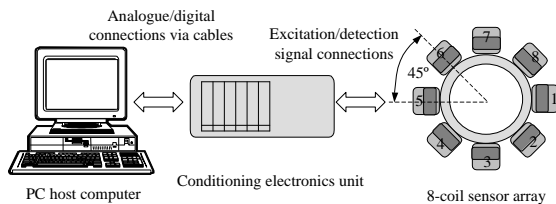


Figure 3 Block diagram of an electromagnetic induction imaging system

An industrial case study of this system was to image molten steel flow profiles. In the continuous casting of steel, control of molten steel delivery through the submerged entry nozzle (SEN), between tundish and mould, is critical to create the optimum laminar flow patterns. These factors influence the surface quality and cleanliness of steel. Typical examples of steel flow patterns within nozzle are full, half and annular flow with the possible transition between these flow patterns during casting depending on the casting conditions. The imaging system has been successfully used to visualize molten steel flow profiles using real data acquired during continuous casting at Corus [6].

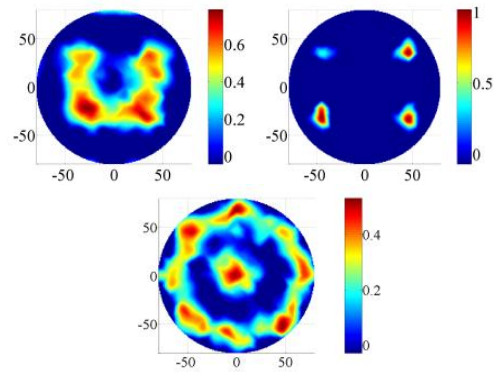


Figure 4 Image examples, (left) a u-shape object, (middle) three 19 mm copper rod and one 12.5 mm aluminium rod at square positions and (right) a ring with copper rod inside

### III. CONDITION MONITORING

#### A. Insulation Condition Monitoring

Failure of the solid insulating systems of high voltage equipment (rotating machines, transformers, bushings etc.) generally results in partial discharges (PD) being generated (small electrical sparks within the insulation). Once prevalent, partial discharges then represent the dominant mechanism of degradation. It can be therefore appreciated why the detection and characterisation of partial discharge activity is a key approach in insulation system condition monitoring. PD condition monitoring of high voltage equipment would indicate the state of health of power equipment and hence determine whether and when refurbishment or replacement is to be arranged.

A digital partial discharge detection system requires data acquisition, storage and computer-based digital processing of PD signals, which occur in the form of individual or series of electrical pulses. By acquiring the data at a convenient, non-intrusive location, usually the phase or neutral terminals of the power equipment under test, statistical quantities such as the PD magnitude, the time of occurrence of discharge with regards to the ac power cycle and their variation with time can be obtained. Furthermore, based on these quantities, the mechanisms of PD activity can be categorised, and the nature, form and the extent of degradation can be inferred.

However, the analysis of PD measurements in the field is often hampered by electrical noise, which necessitates the use of sophisticated signal processing techniques. As a

powerful mathematical means in signal processing and analysis, the wavelet transform was developed to decompose/reconstruct a given signal in varying scales and characterise the signal in both the time and frequency domain simultaneously with variable resolution. For this reason, it is particularly suitable for the analysis of transient, irregular and non-periodic signals such as in the case of PD signals [7, 8]. Fig. 5 gives the denoising results, as an example, of a practical PD data set. It is evident that PD signals at levels comparative to and below electronic noise can be fully extracted.

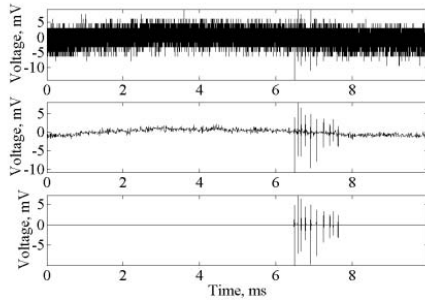


Figure 5 Raw data, denoised signal, and the extracted PD pulses

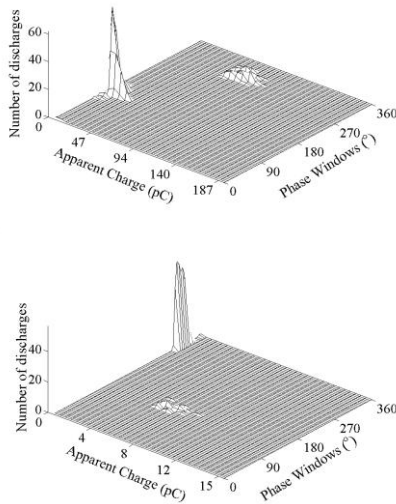


Figure 6 PD patterns of a cable under different discharge geometries

The so-called  $\phi$ - $q$ - $n$  patterns have been widely utilised to identify the origin of the PD source [9]. The  $\phi$ - $q$ - $n$  pattern is created by first dividing a full ac cycle into a number of small phase windows, then calculating the magnitude of the discharges  $q$  and number of the discharges  $n$  with respect to each phase window  $\phi$ , and finally displaying  $n$  as a function of  $\phi$  and  $q$ . These fingerprints are statistical in nature and evolved from a large number of PD events within an adequate monitoring duration. Fig. 6(top) shows the PD  $\phi$ - $q$ - $n$  pattern of a cable when energised at 3.5 kV. This fingerprint claims explicitly that small amount of discharges with large magnitude appear in the negative half cycle, whereas large number of discharges with small magnitude appear in the positive half cycle for this particular discharge

geometry. Fig. 6(bottom) gives the PD  $\phi$ - $q$ - $n$  pattern generated at a test voltage of 2.5kV with a plane-point discharge configuration. As expected, this PD geometry results in an  $\phi$ - $q$ - $n$  pattern with an opposite discharge distribution to that shown in Fig. 6(top).

### B. Generator Condition Monitoring

A significant improvement in operational reliability of generator can be achieved when traditional off-line inspection methods are supplemented with on-line monitoring systems [10]. With these systems, the more traditional time-based maintenance can be shifted to condition-based maintenance allowing for a significant reduction in maintenance costs and unscheduled outages.

At present, monitoring concentrates on specific areas of the generator. For example, partial discharge is used for assessing the condition of the stator winding insulation. The technique detects changes in PD activity by analysing the PD patterns. The stator end winding vibration monitoring can identify high or deteriorating stator end winding and end winding support vibration levels, thereby reducing the risk of fretting, bar damage related to fatigue failure and insulation breakdown. Through stator water temperature monitoring, temperature changes caused by conductor blockages can be identified, enabling appropriate actions to be taken to prevent damage to stator coils. Likewise, rotor condition monitoring and analysis can be made by monitoring shaft voltage and current and by monitoring rotor flux leakage. The shaft voltage/current monitoring can automatically detect shaft rubs, grounding problems and electro-erosion resulting in bearing pitting. Rotor flux monitoring can automatically detect rotor inter-turn shorts at earliest possible stage and identify fault location and magnitude if detected.

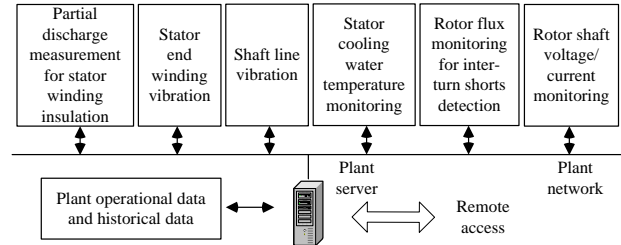


Figure 7 The schematic diagram of a generator condition monitoring system

Fig. 7 shows the architecture of a generator condition monitoring including stator and rotor monitoring data. Interactions between different monitoring modules are taken into account. With this system, the generator's condition can be assessed, which can extend period of safe operation if fault is monitored using online trending capability and ensure correct asset management decisions to be made. However, condition monitoring needs smart technologies to diagnose faults and prognose failures. Recent work [11, 12] has showed that based on historical data, sensor data of generator can be predicted for fault detection and diagnosis by means of system identification, fuzzy logic, and neural networks. Fig. 8 shows the winding temperatures estimated by electrical power and



cooling air temperature by a neural network. By observing the residuals, i.e., by comparing estimated values and measurement data, an adaptive thresholding rule can be defined to evaluate real conditions of the equipment.

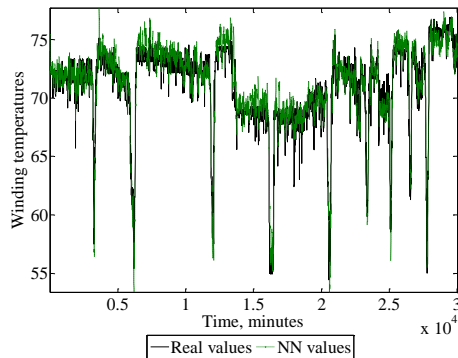


Figure 8 Prediction of sensor data using natural network model

### C. Integrated Control and Monitoring

The health condition changes of control systems will have a significant impact not only on the structural performance of controllers but also on the operational performance and hence the safety of the devices. The monitoring of controllers and the associated sensors, in combination with the time-varying conditions, will reveal the dynamic nature of the signals and therefore can extract valuable features that support integrated control systems.

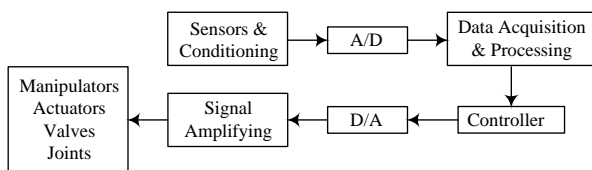


Figure 9 An example of the physical control system

Take a robot as an example. Its control system mainly consists of sensors, A/D converters, data acquisition and processing, central controller, D/A converters, amplifiers, actuators and manipulators (see Fig. 9). The scheme can generally be implemented in two aspects. Firstly, monitoring module accepts signals from control systems and estimates the control outputs such as the positions of manipulators and links/joints by means of model-based methods. By comparing their estimates and real outputs, the fault conditions of control system can be evaluated and identified using adaptive and intelligent approaches. This will provide information to the control system, which responds to the occurrence of a fault ensuring that the faulty system is still well behaved. Therefore condition based control systems capable of integrating condition monitoring technologies and control schemes with greater intelligent advisories would be a challenging topic.

## IV. CONCLUSIONS

The NDT applications have been now extended to cover medical diagnosis, intelligent robotics, on-line

manufacturing process control and security screening. NDT can be an effective tool used for the structural health monitoring. For condition monitoring, there are many examples of successful research and applications to date, including development of built-in smart sensors, high-speed data acquisition, sophisticated monitoring software, remote monitoring, and ICT platforms. These techniques have found a wide range of applications in the power industries from power electronics, rotating machineries through to civil infrastructures [13-16].

There are a few critical challenges for condition monitoring. One of them arises from the nonstationary behaviours of measurement signals. The nonstationary feature of measurement data becomes even more complex if condition monitoring is applied to renewable energy systems, where renewable sources like wind, waves and tidal currents are intermittent and stochastic in nature. Another challenging aspect is to process and transmit large volumes of data to a remote support centre for further analysis, especially for the online condition monitoring cases. For example, for a modern CCGT (Combined Cycle Gas Turbines) station of 4 generator units, about 3000 monitoring points are required to monitor turbines, generators, auxiliaries and other plant parameters. If a sampling rate is set to one minute, around 4.32 M data are acquired per day. Data storage and transfer would become even more problematic for condition monitoring of wind and wave farms of similar capacity of a CCGT station. In such distributed generation systems, each wind or wave turbine essentially works as a small-scale power station, where apart from turbine and generator, the hydraulics/pneumatics, structures and mooring needs to be monitored as well.

As mentioned previously, condition monitoring needs smart technologies to diagnose faults and prognose failures. The smart technologies can be model based methods, model-free data driven methods and artificial intelligence methods. Model based methods can often give good discrimination between faults but at the cost of detailed models required. Model-free data driven methods do not need to consider the system model; instead they are fitting models purely based on measured data by investigating the relationship between measured inputs and outputs. Artificial intelligence type techniques like neural network (NN), fuzzy logic and expert system are particularly good for dynamic nonlinear systems, which characterise and represent quantitative knowledge from historical data by means of neural networks and qualitative knowledge from operation and maintenance experts by means of expert systems. The integration of artificial intelligence methods with an appropriate analytical tool could be a more robust approach for sensor fusion or sensor data characterisation; consequentially can provide a more robust decision making on the equipment conditions.

It is believed that the reliability and availability of future industrial processes and advanced energy systems can be increased by the application of improved eddy current sensing and condition monitoring techniques.

#### ACKNOWLEDGMENT

The work presented in this paper was financially supported by UK Engineering and Physical Sciences Research Council (EPSRC), European Research Fund under the NEST programme, and National Grid. All the co-authors particularly Prof. A. J. Peyton and Prof. C. Zhou are gratefully acknowledged. The challenge aspects of condition monitoring highlighted in the paper will be investigated by financial support under EPSRC Grant (EP/I037326/1).

#### REFERENCES

- [1] X. Ma and A. J. Peyton, "Eddy current measurement of the electrical conductivity and porosity of metal foams", *IEEE Transactions on Instrumentation and Measurement Technology*, Vol. 55, No. 2, April 2006, pp. 570-576.
- [2] X. Ma, A. J. Peyton and Y. Y. Zhao, "Measurement of the electrical conductivity of open-celled aluminium foam using non-contact eddy current techniques", *NDT & E International*, Vol. 38, No. 5, July 2005, pp. 359-367.
- [3] X. Ma, A. J. Peyton and Y. Y. Zhao, "Eddy current measurements of electrical conductivity and magnetic permeability of porous metals", *NDT & E International*, Vol. 39, pp.562-568, 2006.
- [4] X. Ma, A. J. Peyton, R. Binns and S. R. Higson, "Electromagnetic techniques for imaging the cross-section distribution of molten steel flow in the continuous casting nozzle", *IEEE Sensors Journal*, Vol. 5, No. 2, April 2005, pp. 224- 232.
- [5] X. Ma, A. J. Peyton, S. R. Higson, A. Lyons and S. J. Dickinson, "Hardware and software design for an electromagnetic induction tomography (EMT) system applied to high contrast metal process applications", *Measurement Science and Technology*, Vol. 17, 2006, pp. 111-118.
- [6] X. Ma, A. J. Peyton, S. R. Higson and P. Drake, "Development of multiple frequency electromagnetic induction systems for steel flow visualization", *Measurement Science and Technology*, Vol. 19, 2008.
- [7] X. Ma, C. Zhou and I. J. Kemp, "Interpretation of wavelet analysis and its application in partial discharge detection", *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 9, No. 3, June 2002, pp. 446-457.
- [8] X. Ma, C. Zhou and I. J. Kemp, "Automated wavelet selection and thresholding for PD detection", *IEEE Electrical Insulation Magazine*, Vol. 18, No. 2, March/April 2002, pp. 37-47.
- [9] X. Ma, C. Zhou and I. J. Kemp, "Novel computer-based processing system for partial discharge detection and diagnosis", *IEEE Instrumentation and Measurement Technology Conference (IMTC2004)*, Como, Italy, May 18-20, 2004, pp. 684-688.
- [10] P. J. Tavner, "Review of condition monitoring of rotating electrical machines", *IET Electric Power Applications*, 2008, Vol. 2, No. 4, pp. 215 – 247.
- [11] G. Liu, M. D. Aspinall, X. Ma and M. J. Joyce, "An investigation of the digital discrimination of neutrons and  $\gamma$  rays with organic scintillation detectors using an artificial neural network", *Nuclear Instruments and Methods in Physics Research A*, Vol. 607, 2009, pp. 620–628.
- [12] X. Ma, "Online intelligent condition monitoring of electrical machines", *Proceedings of the 24th International Congress on Condition Monitoring and Diagnostics Engineering Management (COMADEM2011)*, Stavanger, Norway, 30th May - 1st June 2011, ISBN: 0-9541307-2-3.
- [13] K. Schroeder, W. Ecke, J. Apitz, E. Lembke and G. Lenschow, "A fibre Bragg grating sensor system monitors operational load in a wind turbine rotor blade", *Measurement Science and Technology*, Vol. 17, 2006, pp. 1167-1172.
- [14] G. Betta, C. Liguori, A. Paolillo and A. Pietrosanto, "A DSP-Based FFT-Analyzer for the fault diagnosis of rotating machine based on vibration analysis", *IEEE Transactions on Instrumentation and Measurement Technology*, Vol. 51, No. 6, 2002, pp. 1316 – 1322.
- [15] J. Campos, "Development in the application of ICT in condition monitoring and maintenance", *Computers & Industrial Engineering*, Vol. 60, 2009, pp. 1-20.
- [16] N. Dominelli, A. Rao and P. Kundur, "Life extension and condition assessment: techniques for an aging utility infrastructure", *IEEE Power and Energy Magazine*, Vol. 4, No. 3, 2006, pp. 24-35.

# A Constraint-based Design Risk Management Tool for Design Collaboration

Jian Ruan, Sheng Feng Qin

School of Engineering and Design, Brunel University  
London, UK

Jian.ruan@brunel.ac.uk; Shengfeng.qin@brunel.ac.uk

**Abstract**— This paper proposed a constraint-based Design Risk Management (DRM) tool for the improvement of design decision-making under a collaborative product environment. This tool is developed by incorporating collaborative design features, risk management process and theory of constraints (TOC). The research is focused on: 1) exploration of various design constraints from all possible design variables, 2) identification of critical design variables as design risk factors for design management and decision-making, and 3) evaluation of design risks as a project risk based on Bayesian theorem. This tool has been prototyped to show that design managers can use this tool to gain overall design risks during a product design process for their decision-making and meanwhile it enables other designers (or in general team members) to involve in this risk management process and provide first-hand information on dependency of risk factors, the possibility of the risk-happening and the severity of the risk factors. This tool realizes the upstream risk management strategy so that the decision-making for risk mitigation has a reliable data/information support.

**Keywords**- design risk management, design constraints, design collaboration.

## I. INTRODUCTION

The global design collaboration is becoming a mainstream to new product development (NPD) on the basis of a multi-disciplinary and distributed environment. However, during the process of collaborative design, risk is rarely mentioned. In particular, due to the complexity of design process and the lack of efficient design decision-making mechanism, failures in design collaboration happen quite often across multiple companies (Fuha and Li, 2005). Some design projects cannot deliver the benefits as companies have expected through the collaboration. Moreover, a number of stakeholders, managers and designers expressed their disappointment at not seeing the projected savings in cost and time and critically discredit the value of design collaboration (Bauer, 2002). Thus, there is a clear need to conduct design risk management during the process of collaborative product design to prevent the potential hazards involved and ensure the success of the design project.

Many studies in academia and commercial cases have suggested that risk assessment can be applied as an effective means in the field of design industry. Nevertheless, few of them conducted risk management research associated with design constraints under a collaborative design environment from both theoretical and practical perspectives. In current risk management

practice, many risk management practitioners are design managers, no involvement of front-line designers. This downstream management strategy doesn't invite collaboration on risk management within a collaborative design team. This might subsequently give rise to confusion with excessive discussions within a management team. Therefore, to prevent the failure of design collaboration, it is important to perform a satisfactory risk management on the basis of a heuristic process from an upstream perspective.

The TOC has been broadly acknowledged as a management philosophy, which intends to initiate and implement significant improvement for achieving a higher level of performance (Simatupang, et al., 2004). It is also regarded as a systemic problem-structuring and problem-solving methodology which can be used to develop solutions with both intuitive power and analytical capability in any environment (Mabin and Balderstone, 2003; Rong et al., 2006). In particular, TOC emphasises the cross-functional and interdependent nature of organisational processes by viewing an organisation as a chain of interdependent functions, processes, departments or resources. These interdependencies are regarded as constraints. Moreover, in any complex organizational process, there are only a few constraints that have a significant and immediate influence on the whole system. Typically, these sorts of design constraints can be regarded as the source of the weakest links or risk factors, which might greatly impact the performance of overall results.

Thus, considering that only a few constraints that have significant and immediate influence on the whole system, it is essential to identify and evaluate these constraints in order to support managers and designers for mapping, measuring and migrating design risk variables in a more accurate and efficient manner.

## II. ETHODOLOGY

In order to explore design constraints for general product design management, literature survey and questionnaire survey are designed as the main research methods that support the development of a constraint-based DRM tool. Moreover, literature survey is also conducted to investigate and compare the existing risk assessment techniques.

### A. Literature survey

With the aim of identifying the potential design constraints and appropriate risk criteria relating to collaborative design, a literature survey is conducted in

the field of TOC, risk theory and design management. This survey is mainly based on E-journal databases, especially Emerald and Science-Direct. These two databases are selected as a major source because of the focusing on both engineering and management research fields. Moreover, several papers are selected as major journal or proceeding sources for the literature survey due to the fact that most related research works are included.

In addition, some academic books are reviewed to discover related design constraints, such as “Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design” (Stapelberg, 2009), “Engineering design: a systematic approach” (Pahl et al., 2007). As a result, 30 design constraints are summarized and categorized into three collaborative levels respectively: task-dependency, role-interaction and resource-integration (Rong et al., 2006; Ruan and Qin, 2009).

### B. Questionnaire Survey

In order to identify critical design variables as design risk factors for design management and decision-making, questionnaire survey is used as a proper method to gain opinions or suggestions from experienced design managers and ordinary designers.

The questionnaire survey attempts to explore the most critical design constraints and how to evaluate a design constraint as a design risk factor and what the risk evaluation criteria are based on the results from the literature survey. These critical design constraints and related risk evaluation criteria have a significant effect on the reliability of risk management. More specifically, the questionnaire strongly focuses on the evaluation of specific design constraints. The DRM tool is mainly relied on the accuracy and reliability of evaluated design constraints and risk criteria.

The questionnaire survey is conducted on the basis of a web-based questionnaire survey system. The survey system allows multi types of questions to be set up in a questionnaire consisting of the close-ended questions, the open-ended questions and the ranking questions. After uploading the questionnaire, it can be sent to the target participants by email. Participants can answer questionnaires on line which the data will be saved automatically in an online database. The main advantage of the web-based questionnaire survey is that the questionnaire can be easily created in terms of research objectives and deliver to targeted participants. Moreover, the data collected from web can be exported as an Excel document and can be used for statistical analysis in a straight manner. In the questionnaire survey, design managers are targeted as the main participants. The contact details of these participants are explored from the directories of design companies and research institutes in terms of internet. 150 invitation emails are sent out with 59 valid feedbacks.

### C. Simulation

The In order to evaluate whether our upstream design risk management strategy and associated risk management method can be implemented as a design risk management tool to support both design managers and ordinary

designers in a collaborative design team to work together to manage their design risks collaboratively, a prototype simulation software has been implemented based on Bayesian theorem.

Simulation for prediction of design risks is used as a substitute for the experimentation and intervention on the actual system when such experimentation is greatly time consuming, costly, and inconvenient. More specifically, simulation can provide researchers with practical feedback before designing real systems. Consequently, the researchers may explore or compare the merits of alternative design decisions during the design stage rather than the manufacturing stage. Thus, the cost of design can be decreased significantly. Moreover, by approaching a project at a higher level of abstraction, simulation can help the researchers for enhancing comprehensive understanding of project's structures and components regardless of its inherent complexity.

In this paper, a simulation prototype is developed for the evaluation of the DRM tool which incorporated with the constraint-based collaborative design features, the generic risk management process and the Bayesian computation method. Unified Modelling Language (UML) and Visual Basic.NET are utilised during the process of the development. More specifically, UML is applied in order to generate an UML User Case Diagram, which provides a holistic guide for general software development. Visual Basic.NET is employed for the creation of a structured graphic interface and an interactive information flow for the DRM tool.

## III. RESULTS

### A. Design Constraints

Existing literature on design constraints are not only produced from a technological aspect, but also generated from a management perspective (Wang et al., 2006; Rong et al., 2006; Pahl et al., 2007; Stapelberg, 2009). Some of these constraints have a profound influence on the performance and results of collaborative design projects. In the context of product design, constraints can be generally described as restrictions and limitations that constrain the implementation of a design project in NPD. Designers and managers should consider a multitude of schedule, technical, financial, social, environmental, and political constraints during the process of design projects (Wang et al., 2006; Rong et al., 2006; Pahl et al., 2007). More importantly, the constraints are identified by considering the effects of the failure of each identified performance variable in the field of product design (Stapelberg, 2009). In other words, design constraints can be used to guard against failure or restrict the design specifications and requirements of the design space. As De Mozota (2003) suggested, every problem posed to a designer demands that the constraints of technology, ergonomics, production and the marketplace be factored in and a balance be achieved. The finest performance with respect to relative constraints would have higher safety margin, which give rise to more reliable product design (Rong, et al., 2006; Stapelberg, 2009).

Thus, the feasible design solutions with no violated constraints result in an acceptable or satisfied outcome. Design is a process that balances needs and functional requirements against various constraints such as material, technological, economical, physical, functional, operational, environmental, legal, and ergonomic factors (Pahl et al., 2007; Stapelberg, 2009). Thus, the design is a cross-functional process that integrates the constraints from all aspects of design management. In order to increase the odds of the success, collaboration and management among these constraints are highly recommended (Rong et al., 2006; Pahl et al., 2007; Stapelberg, 2009). In addition, design constraints must be identified as early as possible because they limit the options available to product designers. Therefore, it is crucial to identify the critical design constraints in the earlier stage collaborative design in order to meet the success of NPD.

### B. Bayesian Computation Method

In current literature, several studies suggest that Bayesian theorem can be valuable when applied during the process of design risk analysis (Ferson, 2005). First, Bayesian theorem provides a mathematical framework for performing inference, or reasoning, using probability, which certainly presents an important way to select inputs for a risk analysis (Ferson, 2005). Second, by redefining probability as a subjective quantity rather than a measure of limiting frequencies, Bayesian theorem can compute “credibility intervals” to characterise the uncertainty about parameter estimates. Third, the Bayesian approach might be useful in addressing complex and complicated issues under the circumstances that are similar to the process of collaborative design: 1) there may be little or even no empirical data available for some variables, 2) it may be necessary to employ subjective information from the analyst’s judgment or expert opinion, and 3) uncertainty about the mathematical model used in the assessment may be substantial (Ferson, 2005). Thus, Bayesian method for risk analysis overcomes some limitations of the classical approach for parameter and model selection.

### C. Application with DRM

The In collaborative design risk management, after risk variables are identified, their characteristics need to be assessed and evaluated in order to assist further analysis for risk migration. The Bayesian computation method is an extension of probabilistic risk assessment (PRA), which combined estimated probability and Bayesian’ rule that provide a unified approach for DRM. Moreover, the Bayesian method requires incorporating prior beliefs into the estimation process other than simply considering the likelihood density. It applies widely in several ways for the modelling, design and analysis of simulation experiments. In this paper, the Bayesian approach refers to the parameter inference through observations probability of data with Bayesian’ rule. The detailed of the method incorporated with DRM tool is shown below:

- Assumed that consequence resulting from the identified risk variables is known, and denoted as Probability Risk1 (PR1) and Probability Risk2 (PR2) (both value from 0 to 1), where PR1 indicates the corresponding risk magnitude when risk variables occurred, whereas PR2 refers to risk magnitude when risk variables not occurred.
- Design managers and designers need to assign value of estimated probability (Ep) for risk variables in terms of their knowledge, experience and judgments.
- Calculate the estimated risk (Er) by formulation:

$$Er = PR1 * Ep + PR2 * (1 - Ep) \quad (1)$$

- Where the estimated risk (Ep) is that of risk summation involved both situations whether risk variables arises or not.
- According to Bayesian’s rule, Bayesian probability (Bp), or calls posterior probability is computed, given that a risk variable occurred, then

$$Bp = Ep * PR1 / Er \quad (2)$$

- Calculated Bayesian risk (Br):

$$Br = Bp * PR1 + (1 - Bp) * PR2 \quad (3)$$

### D. A constraint-based DRM tool

The integrity of design constraints and Bayesian computation method can be used to constitute a constraint-based DRM tool that ensures the accuracy and objectivity of mapping, measuring and migrating of the appropriate design risk variables with a desired probability value and the assigned consequence rank. These values are inputted primarily based on design experts knowledge and experience. The tool provides the means by which complex collaboration designs can be properly analysed and reviewed. Such analysis and review are conducted not only from an upstream perspective but also with a perspective of combination of design managers, designers, and related stakeholders at an operational level.

### E. DRM Simulation

In order to evaluate whether the DRM tool can be implemented in industry for design risk management, a simulation prototype is developed incorporating with the DRM tool and the Bayesian computation method. The UML and Visual Basic.NET are utilised to create the structured design interface and an interactive information flow for the DRM. The simulation prototype can be applied and facilitated the operation of verification for the DRM implementation in real design companies in a convenient manner of less time consuming, lower cost and no risk.

#### 1) Initial Risk Management Guide

In general, three types of users are designed into the DRM simulation tool as potential user groups. Participators can choose his/ her role when log on the system. After that, the user needs to input his/her personal details in order to build up the system usage profile and



the history database. This information is generally required in this step for the subsequent data tracing and reference. Besides, four design stages are provided for collaborative design space modelling: conceptual design, detailed design, embodiment design and manufacture design.

Subsequently, an interface fulfilled with DRM functions. Based on the proposed DRM tool, in each specific design space, design risk can be mapped and measured in three distinctive collaborative levels correspondingly: task-dependency level, role-interaction level and resource-integration level (see Fig.1). More importantly, in each selected level, the mapping and measuring of design risk variables are entirely compliant with common risk management methodologies and processes which are generally composed of three stages: risk identification (mapping), assessment (measuring), and mitigation.

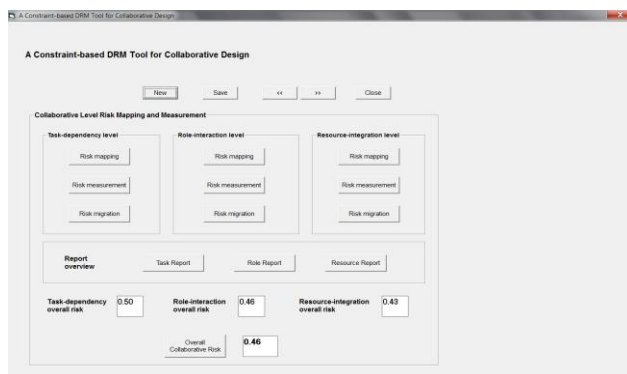


Figure 1. The Main Interface of the DRM Prototype

## 2) Flexible Data Input and Automatic Computation

The DRM software prototype provides flexible data input and automatic computation function with the aid of DRM tool. In risk identification (mapping) stage, participants should represent corresponding design risk variables which can be captured in accordance with evaluated design constraints and risk criteria (see Fig.2). These design risk variables are intended to be demonstrated mainly relating to TOC and encompass a wide range of risk criteria concerns: availability, feasibility, accuracy, reliability and maintainability.

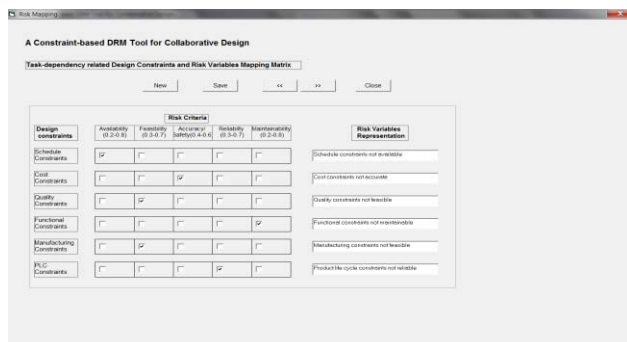


Figure 2. An Illustration of the Input of Risk Variables.

Afterward, in risk assessment (measuring) stage, participants are required to input the assumed probability risk value and to assign the estimated probability value for each identified design risk variable based on their knowledge and experience (see Fig. 3). The DRM will produce the value for its estimated risk, Bayesian probability and Bayesian risk in an automatic manner based on the input data and Bayesian formulae. Moreover, in light of the value of Bayesian risk of each design risk variable, two categorised overall risk values can be calculated. One indicates the risk magnitude of different collaborative levels in the identical design stage. Another one presents overall collaborative risks of different design stages. These values also show in the main interface of the DRM prototype.

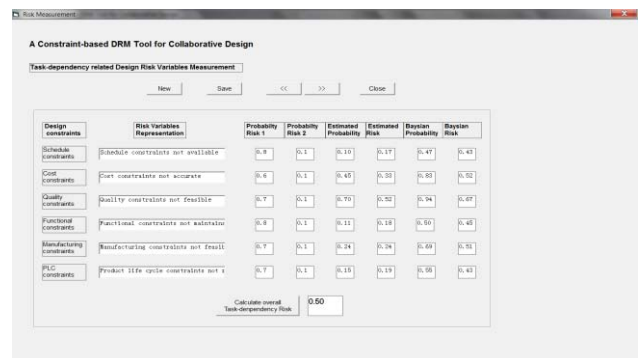


Figure 3. An Illustration of the Input of Estimated Probability and Risk

In the stage of risk migration, users are prompted to provide their own suggestions of risk migration with the purpose of providing a solution for future risk analysis and discussion. For instance, in the task-dependency level, if schedule constraints are not available, participants can reschedule or discuss design project planning at a group meeting, or adopt an alternative schedule. If quality constraints are not feasible, a total quality management (TQM) might need to be considered.

## 3) Multiple Presentations of Results

After the user has completed their DRM data input task, the simulation prototype can compute overall collaborative design risk automatically in three different collaborative levels and calculate overall risk value in four diverse design stages respectively. The results of output can be presented in the form of reports and graphics. Reports show all the detailed information conducted in the process of DRM.

On the one hand, given that risk is associated with likelihood and consequence, a ranking method is used to quantify risk variables levels. The risk migration strategies are recommended for further analysis and discussion. On the other hand, although a variety of risk variables can be identified and assessed on the basis of DRM tool and Bayesian theorem, but overall project risk is not a simple sum of them. Several special design constraints may include the same decision dimensions which might cause duplication in the set of all design risk variables. Thus, it is necessary to identifies and removes the replications, and

clarifies all the risk variables by examining reports. Users could analyse risk variables based on the identification of ranking risk magnitude and the elimination of iterative items.

Besides, two types of graphic diagrams are also provided in the simulated prototype. One of graphic diagrams presents the comparison of overall risk values of different collaborative levels in each design stages for a single user. Another graphic diagram shows the comparison of overall risk values of all users for each design stages. With the aid of these print-out reports and graphic diagrams, users can visualise and contrast the results from different participants straightforwardly.

#### IV. DISCUSSION

This research is similar to a proposal of the integrated risk management in industrial design, which is funded by the European Research by Cooperation Work Program ([www.euresearch.ch/.../Callfiche\\_NMP\\_221206\\_01\\_SM.pdf](http://www.euresearch.ch/.../Callfiche_NMP_221206_01_SM.pdf)). As the value chain based collaborative design and production activities have become more complex with more interrelations and interdependencies, and with new technologies and materials that introduce new risks into a distributed and multidisciplinary environment, the development of integrated approaches and solutions for risk assessment and management is required. This paper focuses on addressing the complexity and reduces the overall risk and impact from a constraint-based perspective. The consideration of social, technical, ergonomics, organisational, financial and environmental factors is adopted at an upstream level during the process of the design risk management.

#### V. CONCLUSION

This paper developed a constraint-based DRM tool for the improvement of design decision-making under a collaborative product environment. This tool is incorporated collaborative design features, risk management process and TOC. The research provides a method to map and measure collaborative design risks from a constraint perspective. Moreover, a Bayesian computation method is incorporated in order to support the DRM tool for the measurement of collaborative design risk in a more efficient manner. In addition, a visual-based

simulation prototype is created in an attempt to further case study evaluation.

#### ACKNOWLEDGMENT

The paper is fully supported by the CAD and Design group of School of Engineering and Design in Brunel University, UK.

#### REFERENCES

- 1) J. Ruan, and S. F. Qin, "A Generic Conceptual Model for Risk Analysis in a Multi-agent Based Collaborative Design Environment", *Proceedings of the 19th CIRP Design Conference – Competitive Design*. Cranfield University, Cranfield, UK March 30-31, 2009.
- 2) J.Y.H. Fuha, and W.D. Li, "Advances in collaborative CAD: the-state-of-the art", *Computer-Aided Design*, Vol 37, pp.571–581, 2005.
- 3) Z.J. Rong, P.G. Li, X.Y. Shao, and K. S. Chen, "Constraint-based Collaborative Design", *IEEE Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design*, 2006.
- 4) W. M. Wang, J.Hu, J. L. Yin and Y. H. Peng, "Uncertainty Management in the Concurrent and Collaborative Design based on Generalized Dynamic Constraints Network (GDCN)", *IEEE Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design*, 2006.
- 5) R.F. Stapelberg, "Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design". London: Springer, 2009.
- 6) G. Pahl, W. Beitz, J. Feldhusen, and H. K. Grote, "Engineering Design: A Systematic Approach". London: Springer, 2007.
- 7) S. Ferson, "Bayesian methods in risk assessment", Report for Service Environnement & Procédés, 2005, Available at: [www.ramas.com/bayes.pdf](http://www.ramas.com/bayes.pdf) [Accessed 10 Nov 2009]
- 8) M. Bauer, "Collaborative Product Design: A Balanced Approach for Enhanced Client Satisfaction and Profitability", 2002. Available at: [http://www.csc.com/aerospace\\_defense/offerings/16309/43118collaborative\\_product\\_design](http://www.csc.com/aerospace_defense/offerings/16309/43118collaborative_product_design). [Accessed 13th June 2009].
- 9) Y.Qiu, Y. and P.Ge, "A risk-based global coordination system in a distributed product development environment for collaborative design, part 1", *Concurrent Engineering*, vol. 15 no. 4, pp.357-367, 2007.
- 10) B. De Mozota, "Design management: using design to build brand value and corporate innovation". NY: Allworth Press, 2003.

# Mixed integer linear programming models for scheduling the LED planting operation on PCBs

Jiaxiang Luo<sup>1,2</sup>, Jiyin Liu<sup>2</sup>

<sup>1</sup>College of Automation Science and Engineering  
South China University of Technology,  
Guangzhou, China,

<sup>2</sup>Business School and Economics  
Loughborough University,  
Leicestershire, UK

**Abstract** — This paper deals with a scheduling problem arising in printed circuit board (PCB) assembly. In this problem, LED components are to be assembled in batches to specified positions on PCBs by a high speed assembly machine and the position sequence for component assembly needs to be optimized. Three different mixed integer linear programming (MILP) models are proposed for the problem. Problem instances based on real data are generated and used to test and compare the models. The models are solved using a commercial software package. The results show that the model with the minimum number of constraints can find the optimal solution in the shortest time.

**Keywords** - Printed circuit board assembly; Component scheduling; MILP models

## I. INTRODUCTION

In PCB assembly lines, components are assembled on PCBs. Although most components can be assembled using Surface Mounting Technology (SMT), some special components such as LEDs can only be assembled using Plated-Through-Hole (PTH) technology. LED components are planted on PCBs by an improved hybrid assembly machine. The machine uses an arm with several heads to carry LEDs and insert them into holes on the PCB and uses a special cutting tool to fold the legs of LEDs at the back of the PCB. To improve the production rate of the machine, optimization of the LED plantation is considered, in which the sequence of the positions on a PCB to plant LEDs and the assignment of head for each position are determined to minimize the total time needed for assembling all LEDs.

This problem was seldom focused on in the literature since the PTH technology was not popular during the past decades. A similar optimization problem is the optimization of mounting SMCs (Surface Mounting Components) through SMT, which has been widely studied. In both problems, components are picked up from feeders and then planted or mounted to the specified positions on a PCB. In the LED plantation problem, LED is the only component type, the components are supplied by a fixed feeder. However, the SMT problem considers more complex types of SMCs. The SMT problem has been extensively studied and many results were obtained. Ayob and Kendall [1] provided a survey of research work on the topic. This optimization problem is often divided into feeder setup optimization and component pick-and-place sequence optimization since the types of the components to be mounted are different. However, even for the component sequence optimization with fixed

feeders, the component picking sequence should be considered for it greatly affects the assembly completion time, while it has no effect on the LED plantation problem. Besides, most of the research results relied heavily on the machine characteristic [2]. So, it is difficult to apply the approaches used for the SMT problem to the LED plantation problem, but the models and approaches provide references.

The formulated models for SMT are dependent on the machine characteristics. The feeder setup optimization and sequence placement problem were formulated as Quadratic Assignment Problem (QAP) and Travelling Salesman Problem (TSP) for the case that only one component could be picked at a time [3-5]. Altinkemer *et al.* considered a more complex situation where the machine had a rotary head and a certain number of components could be picked and mounted [6]. They formulated the problem as a Vehicle Routing Problem (VRP). Ho and Ji proposed IP models for the feeder arrangement and component sequencing where only one component is picked at a time [7]. Ashayeri and Sotirov developed a multi-objective MIP model for placement optimization problem based on batches of components [8]. Though mathematical modelling is a powerful tool, not much work has been done from this aspect.

The main motivation here is to decrease the completion time of each PCB on a hybrid assembly machine. As an effort in pursuing high quality solution, three MILP models are developed for the problem considered in this paper. One is to directly optimize the sequence of positions and head assignment; the other two optimize the sequence of position batches and the sequence of positions and head assignment in each batch. To compare the models, different types of instances are solved to optimum by an optimizer using the three models. The results show that the model with the minimum number of constraints performs best.

## II. DESCRIPTION OF THE ASSEMBLY PROCESS

The hybrid assembly machine is specialized for assembling large number of LEDs as well as other different types of SMCs on a PCB. This paper focuses on the assembly of LEDs. This machine structure is shown in Figure 1. It is equipped with a fixed PCB table, a fixed LED feeder, a movable robot arm with H heads in a line, and a cutting tool under the PCB table. The process of assembling LEDs is as follows.

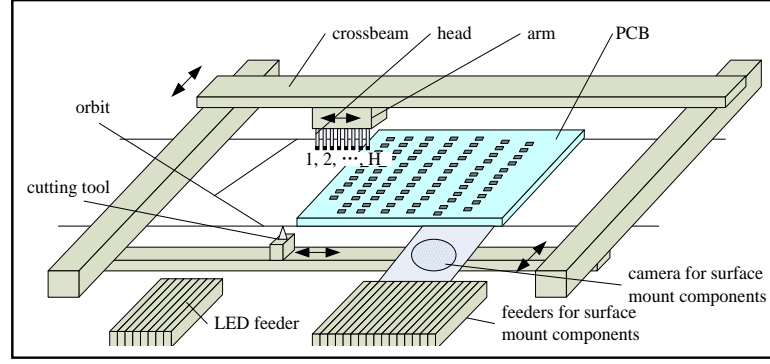


Figure 1. Schematic diagram of a hybrid assembly machine

- 1) The PCB is transferred to a specified position in the machine;
- 2) The robot arm moves to the fixed feeder and pick  $H$  LEDs simultaneously;
- 3) The robot arm moves to a pointed position and plants a LED carried by one of the heads to the position. Synchronously, the cutting tool moves to the position and folds the legs of the LED. Repeat this process until all the  $H$  LEDs carried are planted;
- 4) The arm returns to the feeder to pick up another batch of  $H$  LEDs while the cutting tool moves to the next position. Return to process 3) until all the positions are assembled with LEDs.

The critical decisions for optimizing the process are the position assembly sequencing and head assignment. A round trip of the robot arm picking and planting  $H$  components is called a *repeat*. The assembly completion time composes of assembly processing time and travel time. The former one includes the time planting LEDs and folding the feet, which is almost fixed and needs not be considered. The latter one is the sum of the travel time between each pair of adjacent positions. From position  $i$  to position  $j$ , the travel time  $t(i,j) = \max\{\tau_{ij}, \tau_{ijlh}\}$ , that is, the larger of the time  $\tau_{ij}$  travelled by the cutting tool and the time  $\tau_{ijlh}$  travelled by the arm where the component for position  $i$  is assumed on head  $l$  and that for  $j$  on head  $h$ . Let  $i(x)$  and  $i(y)$  denote the x-coordinate and y-coordinate respectively for position  $i$  in the Cartesian coordinate system. The time  $\tau_{ij}$  and  $\tau_{ijlh}$  can be calculated as follows.

- 1) The time  $\tau_{ij}$  travelled by the cutting tool is proportional to the Chebyshev distance from position  $i$  to  $j$ . Let  $\alpha$  denote the proportion parameter. The time is presented as:
$$\tau_{ij} = \alpha \max\{|i(x) - j(x)|, |i(y) - j(y)|\} \quad (1)$$
- 2) The time  $\tau_{ijlh}$  travelled by the arm from position  $i$  to  $j$  depends on the situation.
  - a. The arm directly moves from position  $i$  to  $j$  if position  $i$  and  $j$  are in one repeat.  $\tau_{ijlh}$  is proportional to the Chebyshev distance from the position of the head carrying the LED for position  $j$  to position  $j$  itself. Let  $\beta$  denote the

proportion parameter and  $\Delta h$  denote the distance between two adjacent heads. the time can be presented as:

$$\tau_{ijlh} = \beta \max\{|i(x) + (h-l) \cdot \Delta h - j(x)|, |i(y) - j(y)|\} \quad (2)$$

- b. The arm returns to the feeder to pick up LEDs before it travels to position  $j$  if positions  $i$  and  $j$  are not in the same repeat.  $\tau_{ijlh}$  is proportional to the Chebyshev distance from position  $i$  to the feeder plus that from the feeder to position  $j$ . Let  $o$  denote the position of the first head when pick up components from LED feeder. The time can be presented as:

$$\tau_{ijlh} = \beta f_{il} + \beta f_{jh} = \beta \max\{|i(x) - (l-1) \cdot \Delta h - o(x)|, |i(y) - o(y)|\} + \beta \max\{|j(x) - (h-1) \cdot \Delta h - o(x)|, |j(y) - o(y)|\} \quad (3)$$

It can be seen that  $\tau_{ijlh}$  varies with the head assignment. Generally,  $\tau_{ijlh} \neq \tau_{ijhl}$  ( $h \neq l$ ). It can also be seen that  $\tau_{ijlh}$  is the minimum if  $h-l=1$  when  $i(x) < j(x)$  and  $l-h=1$  when  $i(x) > j(x)$ . However, such a case will not always happen in a repeat as the x-coordinate of the positions in a repeat are not always in non-decreasing order or in non-increasing order. It is obvious that the completion time is dependent on position-sequence and head-assignment. So, the task of the optimization problem is to determine the position sequence and head assignment. In the following section the problem is modelled as a VRP or a TSP.

### III. MODELLING THE LED ASSEMBLY PROBLEM AND MILP MODELS

#### A. Modelling the LED assembly problem

The problem can be viewed in a direct network, where the nodes correspond to the positions on the PCB and the feeder position. There are three types of operations (see Figure 2): 1) the arm picks up  $H$  LEDs at a time, 2) the arm plant LED one by one to the positions; 3) the cutting tool folds the legs of a LED once it is planted into a position. So, the problem can be described in a direct network  $(V, E)$  where  $V$  is a set of nodes and  $E$  the set of arcs.

Consider the feeder as a deposit, the arm as a vehicle, components as items and positions on the PCB as customers. The problem could be viewed as a VRP: a

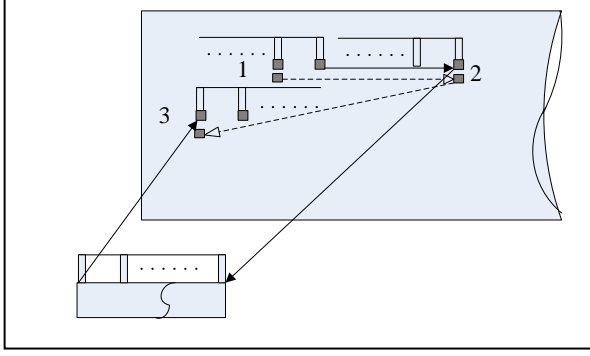


Figure 2. The three types of operations.

vehicle with capacity of  $H$  items starts from the depot, and delivers items to customers, each has a demand of 1 item. Once the vehicle becomes empty, it returns to the depot and picks up another batch of  $H$  items and continue the delivery process until every customer gets one item. The objective is to minimize the total travel time in the whole process. The problem could also be viewed as a TSP problem with city batches, in which each city node denotes a position on the PCB and a batch corresponds to a repeat. The objective is to minimize the travel time.

#### B. Model M1

The idea of this model is to determine the position sequence and head assignment directly. The problem is taken as a TSP. Assume that the heads are indexed from left to right and the cutting tool is always prepared at the first position. This assumption also applies for the next two models. The mathematical model is formulated as follows.

Parameters:

$H$ = Number of heads on the arm.

$N$ = Number of positions (assume  $N$  is a multiple of  $H$  and denote  $T=N/H$ ).

$\tau_{ij}$ = Travel time of the cutting tool from position  $i$  to  $j$ .

$\tau_{ijlh}$ =Travel time of the arm from position  $i$  to  $j$  while the component for position  $i$  is located on head  $l$  and that for  $j$  on  $h$ .

$f_{il}$  = Travel time of the arm from position  $i$  to the feeder while the component for position  $i$  is located on head  $l$ .

$M$ = A large number.

Variables:

$$x_{ik} = \begin{cases} 1, & \text{if the component is planted in} \\ & \text{position } i \text{ at step } k \ (k = 1, 2, \dots, N) \\ 0, & \text{otherwise} \end{cases}$$

$$y_{iql} = \begin{cases} 1, & \text{if the component planted in position} \\ & i \text{ is carried on head } l \text{ in repeat } q \\ 0, & \text{otherwise} \end{cases}$$

$t_{ijk}$  = Time travelled from position  $i$  to  $j$  while the component for position  $i$  is planted at step  $k$  and  $j$  at step  $k+1$ .

$t_0$  = Time travelled by the arm from the feeder to the first position.

$t_T$  = Time travelled by the arm from the last position to the feeder.

Model 1 (M1):

$$\text{Minimize } t_0 + \sum_{k=1}^N \sum_{i=1}^N \sum_{j=1, i \neq j}^N t_{ijk} + t_T \quad (4)$$

S.t.

$$t_{ijk} \geq \tau_{ijlh} + M(y_{iql} + y_{jqh} - 2) + M(x_{ik} + x_{j,k+1} - 2), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T; \\ k=(q-1)H+1, \dots, qH-1; l=1, \dots, H; h=1, \dots, H; l \neq h \quad (5)$$

$$t_{ijk} \geq \sum_{l=1}^H f_{il} y_{iql} + \sum_{h=1}^H f_{jh} y_{jqh} + M(x_{ik} + x_{j,k+1} - 2), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T-1; k=qH \quad (6)$$

$$t_{ijk} \geq \tau_{ij} + M(x_{ik} + x_{j,k+1} - 2), \\ i=1, \dots, N; j=1, \dots, N; k=1, \dots, N; i \neq j \quad (7)$$

$$t_0 \geq \sum_{l=1}^H f_{il} y_{i1l} + M(x_{i1} - 1), i=1, \dots, N \quad (8)$$

$$t_T \geq \sum_{l=1}^H f_{il} y_{iTl} + M(x_{iN} - 1), i=1, \dots, N \quad (9)$$

$$\sum_{i=1}^N x_{ik} = 1, \quad k=1, \dots, N \quad (10)$$

$$\sum_{k=1}^N x_{ik} = 1, \quad i=1, \dots, N \quad (11)$$

$$\sum_{i=1}^N y_{iqh} = 1, \quad q=1, \dots, T; h=1, \dots, H \quad (12)$$

$$\sum_{h=1}^H y_{iqh} = \sum_{k=(q-1)H+1}^{qH} x_{ik}, i=1, \dots, N; q=1, \dots, T \quad (13)$$

The objective (4) is to minimize the assembly completion time. Constrains (5-7) ensure that the time  $t_{ijk} = \max\{\tau_{ij}, \tau_{ijlh}\}$  if the component for position  $i$  is planted at step  $k$  and  $j$  at step  $k+1$ . Constrains (8) and (9) define  $t_0$  and  $t_T$  respectively. Constrains (10-11) force that only one component is planted at each step and the component for position  $i$  is planted at one of the steps. Constrains (12) state that each head in each repeat carries one component for one of the positions. Constrains (13) define the relationships between variables  $y_{iqh}$  and  $x_{ik}$ , which show that the component for position  $i$  is carried by one of the head in repeat  $q$  only if position  $i$  is in this repeat.

#### C. Model M2

The idea of this model is to determine position repeats, repeat sequence, the position sequence and head assignment in each repeat. The problem is taken as a VRP. The mathematical model is formulated as follows.

Variables:

$$x_{iq} = \begin{cases} 1, & \text{if position } i \text{ is in repeat } q \\ 0, & \text{otherwise} \end{cases}$$

$$z_{ik} = \begin{cases} 1, & \text{if the component for position } i \text{ is} \\ & \text{planted at step } k \text{ in its repeat} \\ 0, & \text{otherwise} \end{cases}$$

$$y_{il} = \begin{cases} 1, & \text{if the component for position } i \text{ is} \\ & \text{located on head } l \\ 0, & \text{otherwise} \end{cases}$$

$t_{ijqk}$  = Time travelled from position  $i$  to  $j$  while the component for position  $i$  is planted at step  $k$  and  $j$  at step  $k+1$  in repeat  $q$ .

$t_q$  = Time travelled from the last position in repeat  $q$  to the first position in repeat  $q+1$ .



$t_0$  and  $t_T$  are defined the same as in model 1.

Model 1 (M2):

$$\text{Minimize } t_0 + \sum_{q=1}^T \sum_{k=1}^H \sum_{i=1}^N \sum_{j=1, i \neq j}^N t_{ijqk} + \sum_{q=1}^{T-1} t_q + t_T \quad (14)$$

S.t.

$$t_{ijqk} \geq \tau_{ijlh} + M(y_{il} + y_{jh} - 2) + M(x_{iq} + x_{jq} + z_{ik} + z_{j,k+1} - 4), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T; \\ k=1, \dots, H-1; l=1, \dots, H; h=1, \dots, H \quad (15)$$

$$t_{ijqk} \geq \tau_{ij} + M(x_{iq} + x_{jq} + z_{ik} + z_{j,k+1} - 4), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T; \\ k=1, \dots, H-1 \quad (16)$$

$$t_q \geq \sum_{l=1}^H f_{il} y_{il} + \sum_{h=1}^H f_{jh} y_{jh} + M(x_{iq} + x_{j,q+1} + z_{iH} + z_{j1} - 4), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T-1 \quad (17)$$

$$t_q \geq \tau_{ij} + M(x_{iq} + x_{j,q+1} + z_{iH} + z_{j1} - 4), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T-1 \quad (18)$$

$$t_0 \geq \sum_{l=1}^H f_{il} y_{il} + M(x_{i1} + z_{i1} - 2), i=1, \dots, N \quad (19)$$

$$t_T \geq \sum_{l=1}^H f_{il} y_{il} + M(x_{iT} + z_{iH} - 2), i=1, \dots, N \quad (20)$$

$$\sum_{i=1}^N x_{iq} = H, \quad q=1, \dots, T \quad (21)$$

$$\sum_{q=1}^T x_{iq} = 1, \quad i=1, \dots, N \quad (22)$$

$$\sum_{k=1}^H z_{ik} = 1, \quad i=1, \dots, N \quad (23)$$

$$\sum_{h=1}^H y_{ih} = 1, \quad i=1, \dots, N \quad (24)$$

$$z_{ik} + z_{jk} \leq 3 - (x_{iq} + x_{jq}), i=1, \dots, N; \\ j=1, \dots, N; i \neq j; q=1, \dots, T; k=1, \dots, H \quad (25)$$

$$y_{ih} + y_{jh} \leq 3 - (x_{iq} + x_{jq}), \\ i=1, \dots, N; j=1, \dots, N; i \neq j; q=1, \dots, T; h=1, \dots, H \quad (26)$$

The objective (14) is to minimize the assembly completion time, including the time consumed during each repeat and the time consumed between each pair of adjacent repeats. Constrains (15-17) ensure that both time  $t_{ijqk}$  and  $t_q$  equal to the larger of the time travelled by the arm and by the cutting tool but in different situations. Constraints (19) and (20) define  $t_0$  and  $t_T$  respectively. Constraints (21- 22) force that there are  $H$  positions in a repeat and a position must be in one of the repeats. Constraints (23-24) state that the component for position  $i$  is planted at one of the steps and carried by one of the heads. Constrains (25) and (26) define the relationships between variables. The components for positions  $i$  and  $j$  cannot be carried by the same head or planted at the same step if they are in the same repeat.

#### D. Model M3

The idea of this model is similar to the second one. The difference is that variables for head assignment here are related to repeats. The mathematical model is formulated as follows.

Variables:

$$x_{iq} = \begin{cases} 1, & \text{if position } i \text{ is in repeat } q \\ 0, & \text{otherwise} \end{cases}$$

$$z_{ik} = \begin{cases} 1, & \text{if component for position } i \text{ is planted} \\ & \text{at step } k \text{ in its repeat} \\ 0, & \text{otherwise} \end{cases}$$

$$y_{iql} = \begin{cases} 1, & \text{if the component for position } i \text{ is} \\ & \text{carried by head } l \text{ in repeat } q \\ 0, & \text{otherwise} \end{cases}$$

$t_{ijqk}$ ,  $t_q$ ,  $t_0$  and  $t_T$  are defined the same as in model 2.

Model 1 (M3):

$$\text{Minimize } t_0 + \sum_{q=1}^T \sum_{k=1}^H \sum_{i=1}^N \sum_{j=1, i \neq j}^N t_{ijqk} + \sum_{q=1}^{T-1} t_q + t_T \quad (27)$$

S.t.

$$t_{ijqk} \geq \tau_{ijlh} + M(y_{iql} + y_{jqh} - 2) + M(x_{iq} + x_{jq} + z_{ik} + z_{j,k+1} - 4), \\ i=1, 2, \dots, N; j=1, 2, \dots, N; i \neq j; q=1, 2, \dots, T; \\ k=1, 2, \dots, H-1; l=1, 2, \dots, H; h=1, 2, \dots, H \quad (28)$$

$$t_{ijqk} \geq \tau_{ij} + M(x_{iq} + x_{jq} + z_{ik} + z_{j,k+1} - 4), \\ i=1, 2, \dots, N; j=1, 2, \dots, N; q=1, 2, \dots, T; \\ k=1, 2, \dots, H-1 \quad (29)$$

$$t_q \geq \sum_{l=1}^H f_{il} y_{iql} + \sum_{h=1}^H f_{jh} y_{j,q+1,h} + M(x_{iq} + x_{j,q+1} + z_{iH} + z_{j1} - 4), \\ i=1, 2, \dots, N; j=1, 2, \dots, N; i \neq j; q=1, 2, \dots, T-1 \quad (30)$$

$$t_q \geq \tau_{ij} + M(x_{iq} + x_{j,q+1} + z_{iH} + z_{j1} - 4), \\ i=1, 2, \dots, N; j=1, 2, \dots, N; i \neq j; q=1, 2, \dots, T-1 \quad (31)$$

$$t_0 \geq \sum_{l=1}^H f_{il} y_{i1l} + M(x_{i1} + z_{i1} - 2), i=1, 2, \dots, N \quad (32)$$

$$t_T \geq \sum_{l=1}^H f_{il} y_{iTl} + M(x_{iT} + z_{iH} - 1), i=1, 2, \dots, N \quad (33)$$

$$\sum_{i=1}^N x_{iq} = H, \quad q=1, 2, \dots, T \quad (34)$$

$$\sum_{q=1}^T x_{iq} = 1, \quad i=1, 2, \dots, N \quad (35)$$

$$\sum_{k=1}^H z_{ik} = 1, \quad i=1, 2, \dots, N \quad (36)$$

$$z_{ik} + z_{jk} \leq 3 - (x_{iq} + x_{jq}), i=1, 2, \dots, N; \\ j=1, 2, \dots, N; i \neq j; q=1, 2, \dots, T; h=1, 2, \dots, H \quad (37)$$

$$\sum_{i=1}^N y_{iqh} = 1, \quad q=1, 2, \dots, T; h=1, 2, \dots, H \quad (38)$$

$$\sum_{h=1}^H y_{iqh} = x_{iq}, \quad i=1, 2, \dots, N; q=1, 2, \dots, T \quad (39)$$

The objective (27) is to minimize the assembly completion time. Constrains (28-31) ensure that both time  $t_{ijqk}$  and  $t_q$  equal to the larger of the time travelled by the arm and by the cutting tool but in different situations. Constraints (32) and (33) define  $t_0$  and  $t_T$  respectively. Constraints (34-35) force that there are  $H$  positions in a repeat and a position must in one of the repeats. Constraints (36) state that the component for position  $i$  is plant at one of the steps. Constrains (37) guarantee that the components for position  $i$  and  $j$  could not be planted at the same step if they are in the same repeat. Constraints (38) state that each head in each repeat carries a component for one of the positions. Constrains (39) shows that component for position  $i$  is carried by one of the head in repeat  $q$  only if the position is in this repeat.

#### IV. COMPUTATIONAL RESULTS AND MODEL COMPARISON

To test and compare the models, computational experiments are carried out. The instances are generated according to the production data. In production,  $H=8$  and  $N \geq 8$ . However, for a production instance with  $N=8$  and  $H=8$ , the optimizer did not solve any model to optimum within 6000s in preliminary experiments. So,  $N$  and  $H$  are set smaller than those. However, the position coordinates are selected based on practical data.

Table 1 gives the numbers of variables and constraints for the three models. It shows the following relationships:  $vn_1 > vn_3 > vn_2$  and  $cn_1 < cn_3 < cn_2$ . There is a slight difference between  $vn_1$ ,  $vn_2$  and  $vn_3$  when  $N$  and  $H$  are small. For the

number of constraints, the highest order for all models is 3 and there is a coefficient difference of 1 between them.

Five sets of data, each with 5 instances, are generated. Each problem instance is formulated into the three models and solved by Xpress optimizer. For comparison purposes, the results obtained are presented in Table 2. Based on the results, a number of observations are highlighted below.

- In case 1-10, the problem sizes are very small and the running times for the three models do not have much difference. That is because for small problems there is only a slight difference between the numbers of variables and constraints for different models.

TABLE I. COMPARISON MODELS ON RUNNING TIME

Model	Number of variables	Number of constraints
M1	$vn_1 = N^3 + N + 2$	$cn_1 = N^3[(H-1)^2 + 1] + N^2[(T-2) - (H-1)^2H] + (5+T)N$
M2	$vn_2 = N^3 - N^2(T+1) + N(2T+2H) + T + 1$	$cn_2 = N^3[(H-1)^2 + 3] + N^2[(T-5) - (H-1)^2] + (7-T)N + T$
M3	$vn_3 = N^3 - N^2(T+1) + N(T+H) + T + 1$	$cn_3 = N^3[(H-1)^2 + 2] + N^2[(T-4) - (H-1)^2] + (7-T)N + T$

TABLE II. COMPARISON MODELS ON RUNNING TIME

No.	N×H	Objective	Running time(s)			Iterations of finding the optimal solution		
		M1	M1	M2	M3	M1	M2	M3
1	4×4	435.08	0	0	0	2722	3048	2282
2		448.54	0	1	1	2556	3108	2390
3		419.72	0	0	0	1357	2662	1398
4		511.64	0	0	0	2465	2277	1821
5		499.08	1	1	1	2465	2358	60
6	5×5	442.38	6	6	5	7519	11093	7659
7		430.22	6	8	7	9567	9414	9494
8		504.06	2	2	2	5284	6266	5284
9		507.46	6	6	6	7290	8281	8252
10		519.04	15	18	18	3383	6745	7665
11	6×3	864.60	8	11	10	13028	10586	9668
12		840.27	11	15	13	8097	3064	6165
13		992.71	5	14	5	8462	13981	13646
14		1044.52	13	20	15	13202	1853	8647
15		1044.60	8	11	10	19713	15284	5234
16	6×6	451.47	96	102	106	20496	21420	9775
17		447.51	209	196	231	106709	67949	157066
18		528.03	50	65	59	18208	16247	19268
19		514.29	83	93	120	32930	32455	44551
20		507.81	373	403	358	15842	16848	11253
21	8×4	864.60	2351	4179	4257	808222	46170	32998
22		931.36	270	636	1048	197952	439032	1202516
23		977.86	651	867	1341	178785	483085	1337760
24		1043.69*	>6000	>6000	>6000	383879	2363262	611924
25		1024.18	622	1724	2013	8738	344242	5357

- The running time increases with the number of components. When the number of components increases to 8, the running time for some cases becomes unmanageable and it may increase to several hours. For instances with larger sizes, running time may be even longer.
- The running time varies with instances even for the same problem size. For example, the running time for cases 22 and 23 is much smaller than that for the other cases with the same size. That is because the optimizer uses a branch and bound method and depending on the data the quality of lower bounds generated during the search are different.
- The results show that model 1 performs best for the problem studied in this paper especially when the problem size is large. Although it has relatively largest number of variables, for this problem it is not a key factor of determining the computation time compared with the number of constraints. Large number of constraints makes the LP relaxation harder to solve at each node of the branch and bound process, which may prolong the running time.

#### V. CONCLUSIONS

The problem of scheduling LED plantation on a high-speed assembly machine is an important decision problem to improve the production rate of the machine. Three different MILP models are proposed for it with the objective of minimizing the time travelled by the arm and the cutting tool of the machine to complete the LED assembly task. The models are solved using the Xpress optimizer. The numerical results indicated the first model with minimum number of constraints is most efficient. Further work will focus on finding a near optimal solution in very short time for large sized problem.

#### ACKNOWLEDGMENT

This research was partially supported by National Natural Science Foundation of China (Grant No. 60804053 and 60835001), Specialized Research Found for the Doctoral Program of Higher Education of China (Grant No. 200805611065) and the Fundamental Research Funds for the Central Universities of China.

#### REFERENCES

- [1] M. Ayob, and G. Kendall, "A survey of surface mount device placement machine optimization: Machine classification", *European Journal of Operational Research*, vol. 186, pp. 893-914, 2008.
- [2] W. Ho and P. Ji, "Component scheduling for chip shooter machines: A hybrid genetic algorithm approach", *Computers and Operations Research*, vol. 30, pp. 2175-2189, 2003.
- [3] M. Ball, M. Magazine, "Sequencing of insertions in printed circuit board assembly", *Operations Research*, vol. 36, pp. 192-201, 1988.
- [4] L. F. McGinnis, J. C. Ammons, M. Carlyle, L. Cranmer, G. W. DePuy, K. P. Ellis, C. A. Tovey, and H. Xu, "Automated process planning for printed circuit card assembly", *IIE Transactions*, vol. 24, pp. 18-30, 1992.
- [5] I. Or and E. Demirkol, "Optimization issues in automated production of printed circuit boards: Operations sequencing and feeder configuration problems", *Transactions on Operational Research*, vol. 8, pp. 9-23, 1996.
- [6] K. Altinkemer, B. Kazaz, M. Köksalan, and H. Moskowitz, "Optimization of Printed board manufacturing: Integrated modelling and algorithms", *European Journal of Operational Research*, vol. 124, pp. 409-421, 2000.
- [7] W. Ho and P. Ji, "An integrated scheduling problem of PCB components on sequential pick-and-place machines: Mathematical models and heuristic solutions", *Expert Systems with Applications*, vol. 36, pp. 7002-7010, 2009.
- [8] J. Ashayeri, N. Ma and R. Sotirov, "An aggregated optimization model for multi-head SMD placements", *Computers & Industrial Engineering*, vol. 60, pp. 99-105, 2011.

## A NOVEL APPROACH TO MODELLING AND SIMULATION OF THE DYNAMIC BEHAVIOUR OF THE WHEEL-RAIL INTERFACE

Arthur Anyakwo, Crinela Pislaru, Andrew Ball, Fengshou Gu

University of Huddersfield  
Diagnostics Engineering Research Centre  
Queensgate, Huddersfield, HD1 3DH  
arthur.anyakwo@hud.ac.uk

**Abstract**— This paper presents a novel approach to modelling and simulation of the dynamic behaviour of rail-wheel interface. The proposed dynamic wheel-rail contact model comprises wheel-rail geometry and efficient solutions for normal and tangential contact problems. This two-degree of freedom model takes into account the lateral displacement of the wheelset and the yaw angle. Single wheel tread rail contact was considered for all simulations and Kalker's linear theory and heuristic non-linear creep models were employed. The second order differential equations are reduced to first order and the forward velocity of the wheelset is increased until the wheelset becomes unstable. A comprehensive study of the wheelset lateral stability is performed and is relatively easy to use since no mathematical approach is required to estimate the critical velocity of the dynamic wheel-rail contact model. This novel approach to modelling and simulation of the dynamic behaviour of rail-wheel interface will be useful in the development of intelligent infrastructure diagnostic and condition monitoring systems. The automated detection of the state of the track will allow informed decision making on asset management actions – especially in maintenance and renewals activities.

**Keywords:** modelling; simulation; condition monitoring; systems engineering; wheel-rail contact

### NOMENCLATURE

$R_0$  = Nominal rolling radius of the wheel (460mm)  
 $R_l$  = left wheel rolling radius (mm)  
 $R_r$  = Right wheel rolling radius (mm)  
 $R_{rail}$  = Rail radius (79.37 mm)  
 $a$  = Half length of the semi-axes of  $c$   
in the rolling direction (mm)  
 $b$  = Half length of the semi-axis of contact  
patch in the lateral direction (mm).  
 $I_z$  = Moment of Inertia of the wheelset ( $700 \times 10^6 \text{ kg-mm}^2$ )  
 $K_{py}$  = Lateral suspension stiffness ( $3.86 \times 10^3 \text{ N/mm}$ )  
 $K_{px}$  = Longitudinal spring stiffness ( $850 \text{ N/mm}$ )  
 $C_{py}$  = Lateral damper coefficient ( $8 \text{ Ns/mm}$ )  
 $C_{px}$  = Longitudinal damper coefficient ( $100 \text{ Ns/mm}$ )  
 $f_{11}$  = Longitudinal linear creep coefficient ( $8.06 \times 10^6 \text{ N}$ )  
 $f_{22}$  = Lateral linear creep coefficient ( $8.09 \times 10^6 \text{ N}$ )  
 $f_{23}$  = Lateral/spin linear creep coefficient ( $2.2 \times 10^7 \text{ N-mm}$ )  
 $f_{33}$  = Spin linear creep coefficient ( $1.27 \times 10^7 \text{ Nmm}^2$ )  
 $m$  = Mass of the wheelset (1250kg)  
 $W$  = Axle load (110,000N)  
 $v_x$  = longitudinal creepage  
 $v_y$  = lateral creepage  
 $v_{spin}$  = Spin creepage  
 $l_0$  = Half wheel axle length in central position (742.9mm)

$G$  = Shear Modulus of rigidity =  $80 \times 10^3 \text{ MPA}$   
 $C_{11}$  = Longitudinal creep coefficient  
 $C_{22}$  = Lateral creep coefficient  
 $C_{23}$  = Lateral/spin creep coefficient  
 $C_{33}$  = Spin linear coefficient  
 $d$  = Half distance between the two springs (900mm)  
 $l_0$  = Half wheelset axle distance (742.9mm)  
 $\dot{\phi}$  = Roll velocity

### I. INTRODUCTION

The lateral stability of the wheelset affects the dynamic motion of the railway vehicle. This phenomenon depends on the wheel-rail contact model, wheel-rail profile design, hunting, critical velocity and creep contact forces acting on the contact patch. Hertz theory was applied to solve wheel-rail contact problems [1]. Hertz model runs very fast in real time and is thus used in most railway vehicle dynamic simulations. However for rapidly changing contacts with time and with increased normal contact forces, Hertz model is not suitable since in these situations the contact region becomes conformal. Semi-Hertzian method [2-3] was developed to cater for the variations and increase in the normal contact forces acting on the wheel-rail interface. It uses the geometric intersection of two solids in contact region to find out the shape of the contact patch. Kalker, [4] proposed the exact theory of the wheel-rail contact model by developing a robust algorithm called CONTACT. This model requires so much computation power since the contact patch is discretized into stripes before the tangential creep forces are calculated. Finite Element Method (FEM) [5] was used to model the dynamics of the wheel-rail interface. Due to the enormous computational time required to implement FEM methods, it rarely used for railway vehicle dynamic simulations.

The tangential creep forces play a vital role in wheel-rail rolling theory. Carter solved the 2-Dimensional problem of wheel-rail contact rolling theory using a locomotive wheel and a cylindrical rail [6]. He maintained the fact that the tangential creep forces must not exceed the Coulombs maximum limit. Johnson and Vermeulen extended Carter's theory to 3-dimensional case to include the two smooth half rolling surfaces without spin. Carter's model considered only the relationship between the longitudinal creepage and the tangential forces on the contact patch region. Kalker [7] proposed the linear theory for determining the tangential forces acting on the contact patch. A new Heuristic non-linear model [8-9]

developed by Shen for limiting the tangential forces is discussed. Several dynamic models have been developed for wheel-rail interface using the wheel-rail profile geometry. A parametric 3-dimensional wheel-rail contact model was developed to model the dynamics of the wheel-rail contact model [10]. Wickens [11] studied the effect on hunting on a railway vehicle on a straight track. He observed that hunting motion occurs when the critical velocity of the wheelset exceeds the maximum required speed limit of the designed for its operation. Finally a new model was developed to study the dynamic interaction of the wheel-rail contact on a curved track [12]. This study showed that the lateral and longitudinal stiffness has significant effect on critical velocity of the railway vehicle.

In this paper a two dimensional wheel-rail contact model is modeled. Hertz contact model is used to get the contact patch size dimensions and then the Heuristic nonlinear model is applied to limit the creep contact forces. The lateral stability of the wheelset is investigated by solving the system of non-linear equations of the model using Runge-Kutta's method. The lateral stability of the wheelset is then investigated by increasing the forward velocity of the wheelset until it becomes unstable. The proposed model contains; wheel-rail contact geometry, normal and tangential contact problems and equations for describing the dynamic equation (see Fig. 1).

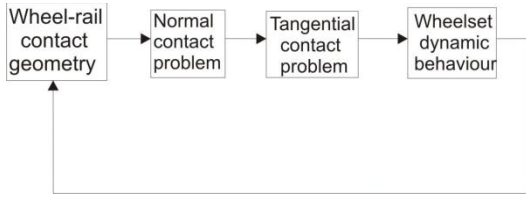


Fig. 1 Dynamic Wheel-rail Contact Model

## II. WHEEL-RAIL CONTACT GEOMETRY

A new conical wheel profile with wheel tread taper 1:20 is used to model the dynamic wheel-rail contact model. A BS 113A rail profile is also used for the model of the rail-profile. The nominal rolling radius  $R_0$  of the wheelset is 460mm while the rail radius  $R_{rail}$  in contact point range is 79.37mm as shown in Fig. 2. This is the contact point range for the wheelset on the track.

Assuming that the yaw angle of the wheelset is very small and can be neglected, the 2-Dimensional model of the wheel-rail contact geometry considering the vertical displacement  $u_z$  and the lateral displacement  $u_y$  is modeled (see Fig. 3).

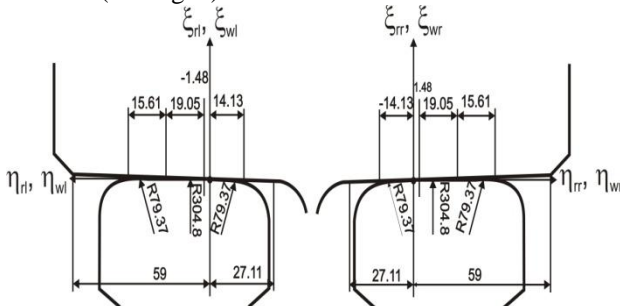


Fig. 2 Wheel-rail contact geometry

The railway track is considered to be rigid and there is no cant applied at both rails. When the wheelset is in central position, the angle made by the horizontal plane is  $\delta_{wr}$  for the right wheel and  $\delta_{wl}$  for the left wheel. Similarly, the co-ordinate of point A at central position with respect to the wheelset frame is  $(l_0, -R_0)$ . When the wheelset is displaced laterally from its central position to the right (as shown in Fig. 3) the rail contact slope formed by the new wheel rail contact point B from A for right wheel profile is  $\delta_{rr}$ . It is a function of the roll angle  $\phi$  and the wheel contact slope  $\delta_{wr}$ . The rolling radius for the right and left wheel tread becomes  $R_r$  and  $R_l$ . The previous wheel-rail contact point on the wheel is now contact point C (see Fig. 3).

The wheel-rail co-ordinates are defined (see Fig. 2) as follows

- $\eta_{wr}$  = Right wheel co-ordinate (lateral direction)
- $\eta_{rr}$  = Right rail co-ordinate (lateral direction)
- $\xi_{wr}$  = Right wheel co-ordinate (vertical direction)
- $\xi_{rr}$  = Right rail co-ordinate (vertical direction)
- $u_z$  = Vertical displacement
- $u_y$  = Lateral displacement

The lateral distance between point A and C is

$$\Delta Y_c = Y_c - l_0 = (u_y + l_0) \cos \phi + (u_z - R_0) \sin \phi - l_0 \quad (1)$$

Similarly, the total vertical distance from point A to C is

$$\Delta Z_c = Z_c - (-R_0) = (u_y + l_0) \sin \phi + (u_z - R_0) \cos \phi + R_0 \quad (2)$$

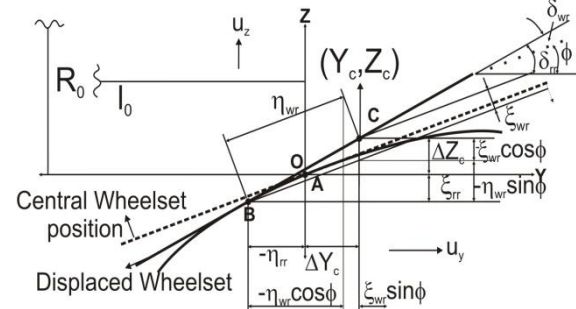


Fig. 3 Right wheel rail geometry

The total lateral distance between point B and C is

$$\eta_{wr} \cos \phi + \xi_{wr} \sin \phi - (\Delta Y_c + \eta_{rr}) = 0 \quad (3)$$

Also, the vertical distance between point B and C is

$$-\eta_{wr} \sin \phi + \xi_{wr} \cos \phi - (\Delta Z_c + \xi_{rr}) = 0 \quad (4)$$

where  $Z_c$  and  $Y_c$  are the vertical and lateral co-ordinates at point C.

For small roll angles  $\cos \phi = 1$ , and  $\sin \phi = 0$ , Eq. 3 and Eqn. 4 simplifies to

$$u_y + R_0 \phi + \xi_{wr} \phi - u_z \phi + \eta_{wr} - \eta_{rr} = 0 \quad (5)$$

$$u_z + l_0 \phi + \eta_{wr} \phi + u_y \phi + \xi_{wr} - \xi_{rr} = 0 \quad (6)$$

The right rail contact slope  $\delta_{rr}$  (see Fig. 3) is

$$\delta_{rr} - \delta_{wr} - \phi \quad (7)$$

Similarly the equations for the left hand wheel-rail contact geometry are

$$u_y + R_0 \phi + \xi_{wl} \phi - u_z \phi + \eta_{rl} - \eta_{wl} = 0 \quad (8)$$



$$u_z - l_0 \phi + u_y \phi - \eta_{wl} \phi + \xi_{rl} - \xi_{wl} = 0 \quad (9)$$

$$\delta_{rl} - \delta_{wl} + \phi \quad (10)$$

Assuming that the contact points are constrained between the region of  $-14.13\text{mm} < \eta_{wr} < 1.48\text{mm}$  and  $-1.48\text{mm} < \eta_{wl} < 14.13\text{mm}$  for the right and left wheel profile, then the wheel profile equations is

$$\xi_{wr} - 0.05\eta_{wr} = 0 \quad (11)$$

$$\xi_{wl} + 0.05\eta_{wl} = 0 \quad (12)$$

The BS 113A rail profile is made up of three main curves with rail radius of 79.37mm, 304.8mm, 79.37 mm (see Fig. 2). The equation of the curves for the region  $-14.13\text{mm} < \eta_{rr} < 1.48\text{mm}$ , right rail contact point region and  $-1.48\text{mm} < \eta_{rl} < 14.13\text{mm}$ , left rail contact point region can be defined as follows;

$$\xi_{rr} = -79.27 + (79.37^2 - (\eta_{rr} - 3.96)^2)^{1/2} \quad (13)$$

$$\xi_{rl} = -79.27 + (79.37^2 - (\eta_{rl} + 3.96)^2)^{1/2} \quad (14)$$

The wheel contact slope is defined as

$$\delta_{wr} = d\xi_{wr} / d\eta_{wr} = 0.05 \quad (15)$$

$$\delta_{wl} = d\xi_{wl} / d\eta_{wl} = -0.05 \quad (16)$$

$$\delta_{rr} = d\xi_{rr} / d\eta_{rr} = (\eta_{rr} - 3.96) / (79.37^2 - (\eta_{rr} - 3.96)^2)^{1/2} \quad (17)$$

$$\delta_{rl} = d\xi_{rl} / d\eta_{rl} = (\eta_{rl} + 3.96) / (79.37^2 - (\eta_{rl} + 3.96)^2)^{1/2} \quad (18)$$

Equations (5) to (18) can be solved synchronously taking  $u_y$  as the input variable using Newton's method which is discussed next.

#### A. Numerical Solution (Newton Raphson Method)

Several methods exist for solving non-linear multi-dimensional equations. The two most common methods include the Newton Raphson's method and the Quasi-Newton method. Newton Raphson method is a numerical method for solving simultaneous non-linear equations. It provides quadratic convergence of the solutions provided the initial conditions are close to the actual solution [15]. The algorithm for implementing this method is

$$x_{k+1} = x_k - J^{-1} f(x_k) \quad (19)$$

where

$J^{-1}$  = Inverse Jacobian matrix of  $f(x_k)$

$x_k$  = initial guess used as the starting point for iterations

$x_{k+1}$  = the new guess

The Newton Raphson algorithm terminates only when the function  $f(x)$  is close to zero. The value of  $x$  at that point is obtained as the solution to the equation. For application to solving the wheel-rail contact geometry equations, Newton Raphson's method is less efficient since for every lateral displacement input  $u_y$ , the initial conditions must be guessed to ensure quick convergence to the solution. A better method for solving these equations is the Quasi-Newton method [15].

#### B. Quasi Newton Method

Identify applicable sponsor/s here. (sponsors)

The Quasi-Newton method is an optimization technique that can be used to solve a system of non-linear differential equations. In Newton-Raphson's method, the Jacobian matrix had to be computed in every iteration but with the Quasi-Newton method, a single Jacobian matrix is determined and thus used for iteration. In Matlab, the function `fsolve` is used to solve a set of simulations non-linear equations using Quasi-Newton's theory of the form

$$f(x) = 0; \quad (20)$$

The algorithms implemented in `fsolve` function are Gauss-Newton method, Levenberg-Marquardt method and the Trust-Region-Reflective method [14]

The syntax used for implementation in Matlab is [14];

$$x = \text{fsolve}(\text{function}, x0, \text{options}) \quad (21)$$

where

$x$  = solution of the equation in vector form

$\text{function}$  = function file containing the set of non-linear simultaneous equations

$x0$  = the initial condition of  $x$

$\text{options}$  = optimization options used for simulations.

Writing the wheel-rail geometry equations into a function file and solving using initial conditions  $x0$  equal to zeros all through, the wheel-rail co-ordinates converged easily to the solution.

For the dynamic wheel-rail contact model, the two most important parameters that are required are the contact angle and the rolling radius difference of the curve.

The contact angle for the left and right wheel-rail geometry defined in Eqn. (7) and (10) would be used for the dynamic model simulation. The rail contact angle plot for the left/right wheel contact is shown below

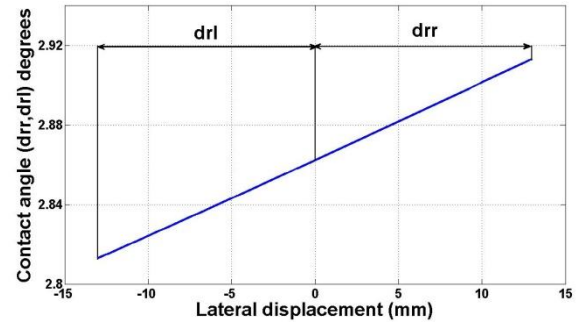


Fig. 5 Contact angle (left and right wheel-rail contact)

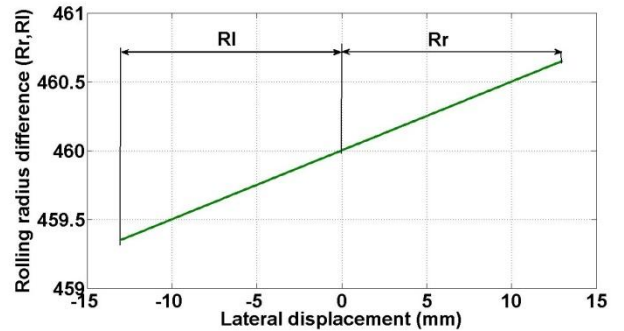


Fig. 6 Rolling radius difference (left and right wheel rail contact)

Fig. 6 shows the Rolling radius difference of the wheelset derived from the left and right vertical wheel co-ordinates  $\xi_{wr}$  and  $\xi_{wl}$  as follows

$$R_r = R_0 + \xi_{wr} \quad (22)$$

$$R_l = R_0 - \xi_{wl} \quad (23)$$

The flangeway clearance for this model is 13mm. Single wheel-rail contact simulations is considered in the wheel tread region for the left and right wheel.

### III THE NORMAL CONTACT PROBLEM

For an applied load on a wheel-rail interface, Normal contact forces develop on the contact patch depending on the total vertical force applied and the contact angle of the wheel-rail contact formed as a result of the lateral displacement  $y$  of the wheelset during motion (see Fig. 7).

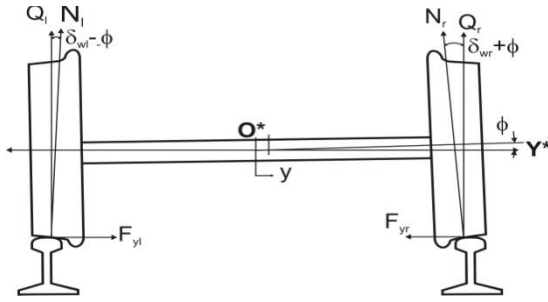


Fig. 7 Vertical and Normal Contact forces acting

The normal contact forces acting on the left and right wheel in static equilibrium is  $Q_r$  and  $Q_l$  given as

$$N_l \cos(\delta_{wl} - \phi) = Q_l \quad (24)$$

$$N_r \cos(\delta_{wr} + \phi) = Q_r \quad (25)$$

The lateral forces  $F_{yr}$  and  $F_{yl}$  can be resolved as follows assuming small roll angles

$$-N_l \sin(\delta_{wl} - \phi) = F_{yl} \quad (26)$$

$$N_r \sin(\delta_{wr} + \phi) = F_{yr} \quad (27)$$

The total lateral force acting on the contact patch is

$$F_{yr} - F_{yl} = Q_l \tan(\delta_{wr} - \phi) - Q_r \tan(\delta_{wr} + \phi) \quad (28)$$

For small contact angles

$$G_r = F_{yr} - F_{yl} = W\phi \quad (29)$$

$G_r$  is the gravitational force. The Gravitational force restores the wheelset back to its central position when it is displaced in the lateral direction.

#### A. Hertz Contact Model

Hertz contact theory predicts the size of the contact patch using the following formulae;

$$ab = mn \left[ \frac{3(1-u^2)}{2E(A+B)} N \right]^{2/3} \quad (30)$$

where  $m$  and  $n$  are the Hertz elliptical constants [2],  $N$  is the normal force(left and right wheel-rail) acting on the contact patch and  $A$  and  $B$  are the relative curvatures given as

$$A = \frac{1}{R}, \quad B = \frac{1}{R_{rail}} \quad (31)$$

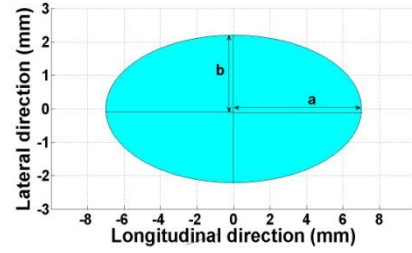


Fig. 8 Elliptical Contact Patch for 0mm lateral displacement (Left/Right wheel-rail contact)

$R$  is the Nominal rolling radius at left and right wheel in the central position or the rolling radii of the left and right wheel as a result of lateral displacement.  $R_{rail}$  is the radius of the rail. In Fig. 8, the Elliptical contact patch for the wheel-rail contact model is shown with values  $a = 7.0045\text{mm}$  and  $b = 2.20245\text{mm}$ . Poisson's ratio ( $u = 0.3$ ) and Young Modulus ( $E = 210000\text{MPa}$ ).

### IV TANGENTIAL CONTACT PROBLEM

The tangential contact problem resolves the tangential creep forces acting on the contact patch. A deviation from pure rolling motion of the wheelset is caused by acceleration, traction, braking and the presence of lateral forces acting on the wheel-rail interface. Creepages are thus formed as a result and can be represented as

$$v_x = \frac{v_{1x} - v_{2x}}{v} \text{ (longitudinal creepage)} \quad (32)$$

$$v_y = \frac{v_{1y} - v_{2y}}{v} \text{ (Lateral creepage)} \quad (33)$$

$$v_{spin} = \frac{\Omega_1 - \Omega_2}{v} \text{ (Spin creepage)} \quad (34)$$

where  $v$  is velocity and  $v_{2x}$ ,  $v_{2y}$ ,  $\Omega_2$  are the real velocities while  $v_{1x}$ ,  $v_{1y}$ ,  $\Omega_1$  are the pure rolling velocities of the wheels in the absence of creep. The longitudinal creepages at (right and left) wheel-rail contact

$$v_{xr} = \frac{v(1 - R_r/R_0) - l_0\dot{\psi}}{v}, v_{xl} = \frac{v(1 - R_l/R_0) + l_0\dot{\psi}}{v} \quad (35)$$

The lateral creepages at the right and left wheel-rail contact

$$v_{yr} = (v^{-1}\dot{y} - \psi)(R_r/R_0) + v^{-1}\dot{\phi} \quad (36)$$

$$v_{yl} = (v^{-1}\dot{y} - \psi)(R_l/R_0) + v^{-1}\dot{\phi} \quad (37)$$

The spin creepages at the right and left wheel-rail contact is

$$v_{spinr} = (v^{-1}\dot{\psi} - \lambda/R_0) \quad (38)$$

$$v_{spintl} = (v^{-1}\dot{\psi} + \lambda/R_0) \quad (39)$$

### A. Kalker's Linear Theory

Kalker established a linear relationship between the developed creepages at the contact patch and the creep forces [7]. The maximum creep forces as determined by Kalker are as follows

Longitudinal creep force

$$F_{xr} = -f_{11}^v v_{xr} \quad (40)$$

$$F_{xl} = -f_{11}^v v_{xl} \quad (41)$$

Lateral creep force

$$F_{yr} = -f_{22}^v v_{yr} - f_{23}^v \text{spinr} \quad (42)$$

$$F_{yl} = -f_{22}^v v_{yl} - f_{23}^v \text{spinl} \quad (43)$$

Spin creep moment

$$M_{zr} = f_{23}^v v_{yr} - f_{33}^v \text{spinr} \quad (44)$$

$$M_{zl} = f_{23}^v v_{yl} - f_{33}^v \text{spinl} \quad (45)$$

where  $f_{11}, f_{22}, f_{23}, f_{33}$  are the linear creep coefficients given computed as

$$f_{11} = GabC_{11} \quad (46)$$

$$f_{22} = GabC_{22} \quad (47)$$

$$f_{23} = G(ab)^{1.5} C_{23} \quad (48)$$

$$f_{33} = G(ab)^2 C_{33} \quad (49)$$

$C_{11}, C_{22}, C_{23}, C_{33}$  are the creep coefficients tabulated by Kalker [6] and  $G$  is the Shear modulus of rigidity of steel.

### B. Heuristic Non-linear Model

The Heuristic non-linear creep model was developed by Shen and White [8] to cater for the non-linearities in the wheel-rail geometry, adhesion limits on the creep force-creepage relationship and the spin creepage effect. The creep forces developed by Kalker's linear theory are limited for high creepages by a saturation constant 'a' developed as follows;

$$a = \begin{cases} \alpha^{-1}(\alpha - 3^{-1}\alpha^2 + 27^{-1}\alpha^3), & \alpha \leq 3 \\ \alpha^{-1}, & \alpha > 3 \end{cases} \quad (50)$$

$$\text{where } \alpha = \frac{\sqrt{(F_{xi})^2 + (F_{yj})^2}}{\mu N} \quad (i = r, l \quad j = r, l) \quad (51)$$

$\alpha$  = unlimited normalized creep ratio

The reduced creep forces now become

$$F_x^r = aF_x \quad (52)$$

$$F_y^r = aF_y \quad (53)$$

$$M_z^r = aM_z \quad (54)$$

### V. WHEELSET DYNAMIC BEHAVIOUR

The dynamic behaviour of the wheelset is studied by summing the total forces acting on the wheelset and then applying Newton's law. In this paper the suspended wheelset is used which includes the primary suspensions

in the longitudinal and lateral direction. The top view of the suspended wheelset is shown below where  $x$  is the rolling direction and  $y$  is the lateral direction.

The suspension forces in the lateral direction and longitudinal direction can be resolved as follows (see Fig. 9)

$$F_{susp} = -2K_{py}y - 2C_{py}\dot{y} \quad (55)$$

$$M_{susp} = -2K_{px}d^2\psi - 2C_{px}d^2\dot{\psi} \quad (56)$$

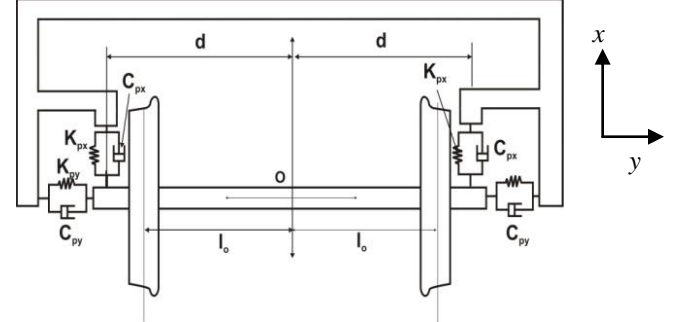


Fig. 9 Top view (Suspended Wheelset diagram)

The equations of motion of the can be derived by combining Eqns. (40 – 45, 55-56) to arrive at the Kalker linear model. For the Heuristic nonlinear model, Eqns. (52-54) is used to replace the maximum creep forces computed in Eqns. (40 – 45). Neglecting the effect of the gyroscopic wheel moment, the two degree of freedom equations of motion comprising of the lateral displacement  $y$ , and the yaw angle  $\psi$  are defined as follows

$$\begin{aligned} m\ddot{y} &= -2f_{22}^v v^{-1} (1 + l_0^{-1} R_0 \lambda + 2C_{py}) \dot{y} - 2f_{33}^v v^{-1} \dot{\psi} \\ &\quad - (2K_{py} + l_0^{-1} W \lambda \lambda) + 2f_{22}^v \psi = 0 \end{aligned} \quad (57)$$

The yaw rotation equation of motion is

$$\begin{aligned} I_z \ddot{\psi} &= 2f_{23}^v v^{-1} (1 + l_0^{-1} R_0 \lambda - I_y \lambda v R_0 l_o) \dot{y} - \\ &\quad - 2(f_{11} + l_o^2 f_{11}) v + 2 * C_{px} d^2 \dot{\psi} - 2f_{11} l_o \lambda / R_o \\ &\quad - (2f_{22} + 2K_{px} d^2 + W \lambda \lambda_o) \psi = 0 \end{aligned} \quad (58)$$

### A. Equations of Motion

The equations of motion of the suspended wheelset Eqn. (50)-(51) is can reduced to a system of first order differential equations;

$$\dot{x}(t) = f[x(t)] \quad (59)$$

where  $x(t)$  is a 4 x 1 state vector variable.

### VI. SIMULATED RESULTS

The ODE45 function in MATLAB implements Runge Kutta 4<sup>th</sup> order method with variable time step for computational efficiency [14]. It solves initial value problems of the form

$$\dot{x}(t) = f(t, x), \quad x(t_0) = x_0 \quad (60)$$

where  $x$  is a state vector of the dependent variables and  $t$  is the independent variable [14].

Applying Ode45 function to solving the equations of motion, the syntax used is

$[t,x] = \text{ode45} (@\text{fun}, \text{tspan}, \text{initialconditions})$   
Where *fun* = function file contain the reduced first order differential equations of motion for the system  
*tspan* = time span for simulation (30 seconds)  
*initialconditions* = initials conditions required for simulation of the dynamic wheel-rail contact model.  
The state variables used for simulation is  
 $x(1)$  = Lateral displacement (Initial condition = 5mm)  
 $x(2)$  = Yaw angle (Initial condition = 0)  
 $x(3)$  = Lateral velocity (Initial condition = 0)  
 $x(4)$  = Yaw velocity (Initial condition = 0)

Further details on the use of Ode45 function can be found in [14]. Fig. 10 and Fig.11 shows Kalker linear Model and the Heuristic Non-linear Model results for various forward velocity inputs. Increasing the forward velocity of the wheelset from 5m/s (5000mm/s) to 40m/s leads to increasing amplitude peaks until the critical velocity is reached. For Kalker's linear model, the critical velocity just before flange contact is 40m/s while for the Heuristic Non-linear model, the critical velocity just before wheel flange contact is 35m/s. It can be readily noted that the critical hunting speeds of the linear model is generally slightly higher than the critical speed for the Heuristic non-linear model. Therefore increase in the forward speed of the wheelset leads to lateral instability and hunting. In most real situations the gravitational forces act as a restoring force to limit these increasing lateral oscillations.

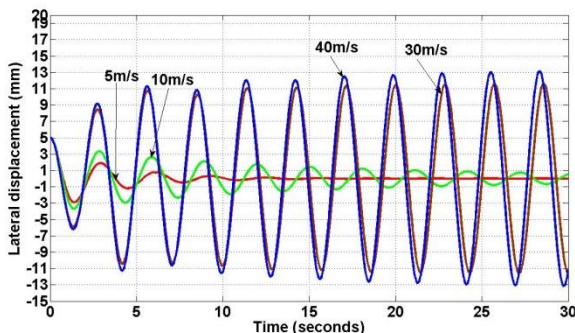


Fig. 10 Lateral displacement of the wheelset for initial velocity 10, 30, 40m/s (Kalker's linear Model)

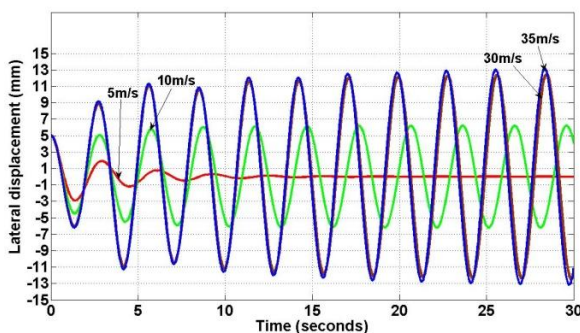


Fig. 11 Lateral displacement of the Wheelset for initial velocity 10, 30, 40m/s (Heuristic Non-linear Model)

## VII

## CONCLUSIONS

In this paper, a new dynamic wheel-rail contact model was developed to study the lateral stability of the wheelset on the track. The equation of motion of a two degree of freedom single suspended wheelset model was derived completely. It was found that as the forward velocity of the wheelset increases, the wheelset becomes unstable on the track due to increasing lateral oscillations. These oscillations are limited by flange contact. This novel approach to modelling and simulation of the dynamic behaviour of rail-wheel interface will be useful in the development of intelligent infrastructure diagnostic and condition monitoring systems.

## REFERENCES

- [1] W. Yan, and F. D. Fischer, "Applicability of Hertz Contact theory to rail-wheel contact problems," *Archive of Applied Mechanics*, vol. 70, May 1999, pp. 255–268.
- [2] J. B. Ayasse, and H. Chollet, "Determination of wheel-rail contact patch in semi-Hertzian Conditions", *Vehicle System Dynamics*, vol. 2, Oxford: Number 2, March 2005, pp. 161 – 172.
- [3] X Quost, M. S. A. Eddahak, J.B. Ayasse, H. Chollet, and P.E. Gautier, "Assessment of the semi-Hertzian method for the determination of the wheel-rail contact patch," *Vehicle System Dynamics*, vol. 44, No. 10, October 2006, pp. 789–814.
- [4] J.J. Kalker, "Rolling contact Phenomena: Linear Elasticity," *CISM International Centre for Mechanical Series*, No. 41, Vol. 6, 2000, 394 pages.
- [5] T. Telliskivi, and U. Olofsson, "Contact mechanics analysis of measured wheel-rail profiles using the finite element method," *Proc. Instn. Mech. Engrs*. Vol. 215, Part F, August 2000.
- [6] J.J. Kalker. "Wheel-rail rolling contact theory," *Wear*, Vol. 144, (1991), pp. 243 – 261.
- [7] J.J. Kalker, "Three dimensional Elastic bodies in rolling contact, Kluwer Academic Publishers, Dordrecht, Boston/London.
- [8] A. Shabana, A., K. Zaazaa, H. Sugiyama, "Railroad Vehicle Dynamics: A Computational Approach," CRC Press, Taylor and Francis Group.
- [9] S. Iwnicki, "Simulation of Wheel-rail contact forces," *Fatigue and fracture of engineering materials and Structures*, Vol. 26, No. 10, 2003, pp. 887-900.
- [10] J. Pombo, J. Ambrosio, M. Silva, "A new wheel-rail contact model for railway dynamics," *Vehicle System dynamics*, Vol. 45, No. 2, February 2007, pp. 165 – 189.
- [11] A. H. Wickens, "The Dynamics of Railway Vehicles Straight Track: Fundamental considerations for lateral stability". *Proceedings, Institute of Mech. Engineers*, Vol. 180, Pt 3F. 1965 pp 1 – 16.
- [12] S. H. Lee, and Y.C. Cheng, "A New Dynamic Model of High Speed Railway Vehicle Moving on Curved Tracks," *Journal of Vibration and Acoustics*, Vol. 130, 2008.
- [13] A. Jaschinski, H. Chollet S.D. Iwnicki, A.H. Wickens and J. Von Würzen, "The Application of Roller Rigs to Railway Vehicle Dynamics," *Vehicle System Dynamics*, Vol. 31, 1999, pp. 345 – 392
- [14] E. Magrab, S. Azaram, B. Balachandran, J. Duncan, and K. Herold, "A Engineers Guide to Matlab," Prentice Hall, 1<sup>st</sup> Ed. August 2000, 512 pages.
- [15] R. Burden, and J. D. Faires, "Numerical Analysis," Thompson Brooks Cole, 8<sup>th</sup> Edition, 2005, 837 pages.



## Time Encoded Signal Processing and Recognition of Incipient Bearing Faults

S. Abdusslam, M. Ahmed, P. Raharjo, F. Gu, A. D. Ball  
University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK  
Corresponding author: [S.Abdusslam@hud.ac.uk](mailto:S.Abdusslam@hud.ac.uk)

**Abstract:** Numerous techniques for rolling bearing monitoring have been presented recently but the challenge lies in finding a reliable and price efficient monitoring system capable of providing an early alarm of bearing defects, thus, the purpose of this paper is to develop a more advanced approach using vibration signal to bearings monitoring based on TESPAP (Time Encoded Signal Processing and Recognition), the results show that TESPAP analysis when applied on its own to raw data vibration signal from different bearing conditions does not produce significant results, however, when combined with envelope signal provides an enhanced and novel method for detection of incipient bearing faults

**Keywords:** Rolling bearing, vibration monitoring, condition monitoring, early alarm, TESPAP analysis, epochs.

### I. INTRODUCTION

The implementation of machine condition monitoring (CM) has been growing significantly to enhance system performance and avert ruinous breakdowns. Numerous new CM methods have been proposed recently in which the key issue has been the application of efficient data analysis methods for precise determination of the machines' condition.

The paper reports a new technique to detect and identify bearing faults using TESPAP singly and in combination. TESPAP distinguishes the shapes of signal waveforms and differentiates between them. This paper reports an experimental investigation of bearing diagnosis using TESPAP with raw time-domain data, envelope analysis of the data and then combining TESPAP with envelope signal. Envelope analysis is particularly useful in extracting fault frequencies from bearings and, while using digital filtering and Fast Fourier Transforms (FFT) which require a large memory to contain all the sampled data, has produced useful results. It was chosen to combine with TESPAP because both techniques identify patterns in the data sets and thus complement each other [1].

This paper assesses the performance of the three methods for a roller bearing with three incipient faults seeded into it. It is shown that, using bearing vibration signals the combination of TESPAP and envelope analysis attains classification results that cannot be reached by either method alone.

### II. TIME ENCODED SIGNAL PROCESSING AND RECOGNITION (TESPAP)

TESPAP is a digital language that originated as a means of coding signals for speech recognition [4]. TESPAP depicts signal waveforms according to its real and complex zeros based on a mathematical waveforms representation which is different from conventional CM techniques.

TESPAP quantisation procedure has been developed to code signals according to the period between two consecutive zero-crossings and the shape of the curve thus contained [6, 7]. This period is named an epoch. Every epoch can be illustrated by two parameters: D (duration - number of samples in the epoch) and S (shape - determined by the number of minima or maxima contained in the epoch). Fig. 1 shows an epoch encoded into its TESPAP parameters where D=17 and S=2.

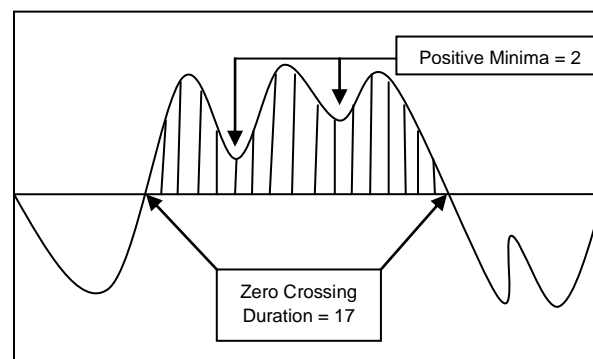


Figure 1. TESPAP single epoch with D=17, S=2

Most signal waveforms can be coded into a limited sequence of numerical descriptors known as the TESPAP symbol stream [8, 9], normally from 1 to 28. In fact 28 symbols have been found to be sufficient to describe most signals adequately. The symbol sequence can be characterised in two ways: A one-dimensional "S-Matrix" vector or two-dimensional vector which is named the "A-Matrix".

The S-matrix can be defined as the TESPAP symbols that record the number of times each TESPAP alphabet symbol occurs in the TESPAP symbol stream, and the A-matrix can be defined as a two-dimensional 28x28 vector matrix that records the number of times each pair of symbols in the alphabet appears n symbols apart in the symbol stream. The A-matrix expresses the



temporal relationship between pairs of symbols [7] and because parameter  $n$  represents the delay between symbols it provides frequency information. Slowly oscillating patterns have  $n > 10$  while higher frequency patterns have  $n < 10$  [10].

In practice for most signal waveforms the TESPAP symbol stream is a limited sequence of numerical descriptors significantly less than 28 [8, 9].

### III. BEARING TEST FACILITIES

To assess the use of TESPAP in bearing fault detection, bearing vibration data was acquired from the bearing test rig shown in Fig. 2. This consists of a 3-phase electrical induction motor combined with a dynamic brake; the stator is free to move so that torque measurements may be taken. The motor is connected to the brake through 3 shafts that are connected by two pairs of matched flexible couplings. These three shafts are held in two bearing housings, one has a cylindrical roller bearing type N406 and the other is a double row self-aligning type 22208 EK. It is the roller bearing that is tested with different faults.

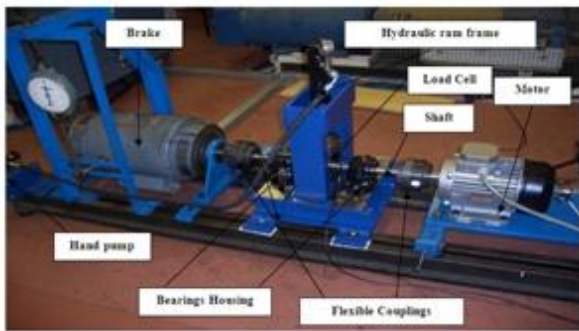


Figure 2. Bearing test rig

Table I lists the specification of the roller bearing. It is a common bearing used for high radial loads. This kind of bearing is convenient for this type of research because different faults can be easily simulated, and each fault has its characteristic frequencies.

Table I. Bearing N406 specification

Elements	Dimension	Characteristic Frequency
Roller diameter	14mm	Outer race = 83.3Hz
Rollers' number	9	Roller element = 48.3Hz
Contact angle	0°	Inner race = 134.4Hz
Pitch diameter	59mm	Cage frequency=9.5Hz

In this paper, a healthy bearing was compared with three identical bearings, each with a fault introduced to outer race, inner race and roller element respectively. The four bearings were tested at shaft rotational speed of 1420 rpm (frequency 23.6Hz) under 50% of torsion load from a DC motor of a maximum 4.0 KW, and 5 bar of radial load that is equivalent to 433 Nm load. The faults were small scratches of 30% of the bearing width and 0.1 mm in depth which was introduced to the outer race, inner race and the roller element of three tested bearings, each bearing is with one fault. These

faults are considered as incipient faults because it causes no influence on the operating performance.

### IV. DATA SETS

Four experiments were performed to acquire data for four bearing conditions: healthy, inner fault, outtrace fault and roller fault. Each experiment acquired data at a sampling rate of 62.5 kHz. The data length for each test was 960,000 points.

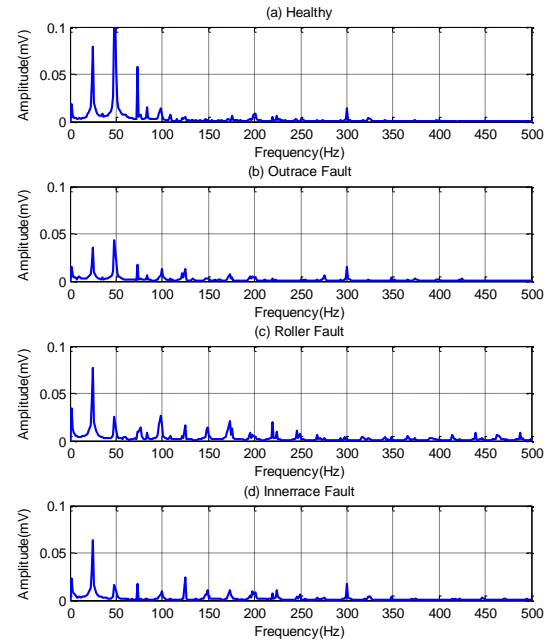


Figure 3. Raw data for a health and three small different faults

To evaluate the quality of the data, commonly used spectrum analysis is performed on the datasets which exhibit little information of the bearing in the time domain because of high noise contamination. Fig. 3 shows the data spectrum of raw data for the healthy and three small different bearings' faults. It shows that the signal is dominated by the shaft frequency at 24.3 Hz and its high order harmonics and it is difficult to identify the bearing feature frequencies from the spectrum. This shows that the bearing signal is very weak components and need to be enhanced so that the performance of TESPAP can be evaluated with good confidence.

It is well know that the most popular method for bearing monitoring is envelope analysis. Fig. 4 shows the envelope spectrum of the datasets in the frequency band from 8kHz to 15kHz. The spectra of the roller and the inner race faults are quite clear as shown in Fig. 4(c) and Fig. 4(d) respectively. The roller and the inner race characteristic faults' frequencies are identified at 48.3 Hz and 134.4 Hz respectively.

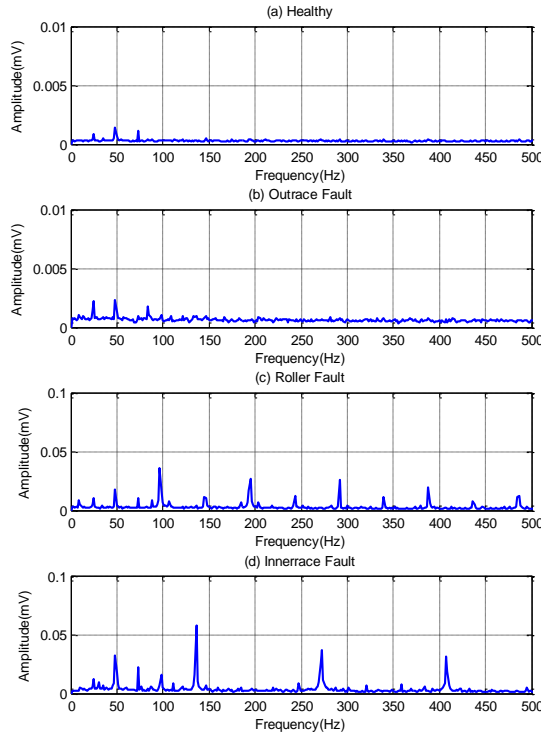


Figure 4. Envelope spectrum for a health and three different faults

When a bearing has no fault it usually has very small vibration amplitudes in the time domain. Moreover, the characteristic frequencies cannot be seen in envelope spectrum. This can be illustrated by the blue solid line in Fig. 4(a). Compared with faulty case in the same plots, the envelope spectrum from the healthy bearing is very flat, i.e. no clear spectral lines can be seen. In Fig. 4(b) the spectra, represents the tiny outer race fault, seems to be smooth and there is insignificant amplitude which cannot be seen easily from the figure.

## V. TESPANALYSIS OF RAW DATA

Having confirmed that the vibration data sets include bearing faults information, they were encoded into their TESPAN symbol streams and then their S and A-Matrices were constructed by a programme written in the Matlab platform. To make comparison between different cases, the Matrices are normalised to the total number of symbols. In total, there are four sets of S-Matrices and four A-Matrices corresponding to healthy, outer race fault, roller fault and inner race fault respectively. In addition, both the raw data and the envelope data are explored with TESPAN to evaluate its noise sensitivity.

### A. TESPAN S-Matrices for raw data

Comparison of the S-Matrices in Fig. 5 shows that there is a difference on TESPAN symbols 2 and 6 that is consistent with fault conditions. In particular, the occurrence rates of symbol 2, 3 and 6 are different between four bearing cases. Based on these differences,

the faulty cases can be identified completely. This means that 3 feature parameters can be used directly to diagnose the faults.

However, the occurrence rates of other symbols do not show a consistent change and not suitable for separating the fault cases.

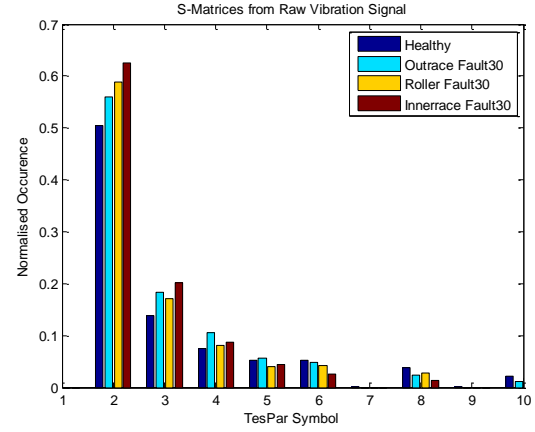


Figure 5. TESPAN S-Matrices for raw data of a healthy and three different faulty bearings

### B. TESPAN A-Matrices for raw data

The A-Matrix represents the waveform in two-dimensions, for the case  $n=2$  the matrix shows the number of pairs two steps apart. The  $n$  attribute is known as the delay between symbols. Here  $n = 2$ , but many other A-Matrices can be formed from the same waveform by changing the value of  $n$ .

Fig. 6 shows the A-Matrices of the same bearing conditions. The patterns for the outer race and the roller faulty bearings appear very similar, where the other patterns are markedly different. From this data set it would be difficult to detect incipient outer race and roller faults using the A-Matrix.

It should be noted however, that the A-Matrix for the inner fault showed quite different pattern for the four bearings which suggests that use of the TESPAN A-Matrices with raw data could be used to detect this particular faults.

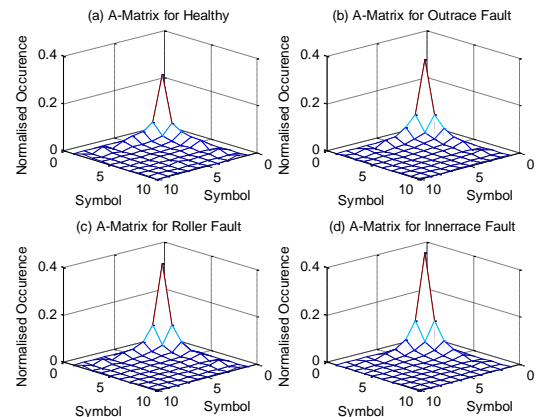


Figure 6. TESPAP A-Matrices for raw data of a healthy and three different faulty bearings

## VI. TESPAP ANALYSIS WITH ENVELOPE SIGNAL

Fig. 4 shows that envelope analysis alone did not show any significant difference between the healthy bearing and the incipient outer race faulty bearing.

### A. TESPAP S-Matrices for envelope signal

Fig. 7 shows the S-Matrices patterns for the combined TESPAP-envelope analysis for the reference bearing and for three different initial fault locations. Both the differences between individual symbol values and between the overall trends for the healthy bearing and the bearing with different incipient faults' locations are highly significant.

Fig. 7 shows unambiguously that the S-Matrix patterns allow clear differentiation of the healthy bearing from the incipient faulty bearings in particular the initial outer race fault which was not detected by envelope analysis independently.

However, it is also possible, based on Fig. 7 to differentiate between the initial bearings fault by using the relative values at  $s = 2, 5, 6$  and  $8$  which show the best features.

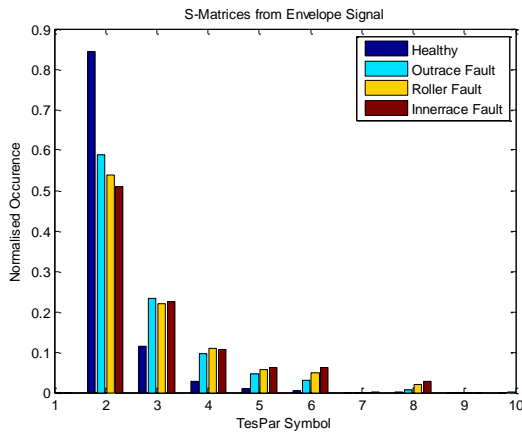


Figure 7. TESPAP S-Matrices for envelope signal of a healthy and three different faulty bearings

### B. TESPAP A-Matrices for envelope signal

The patterns are discernible in the A-Matrices shown in Fig. 8, they provide clear and obvious differences for the separation between the healthy bearing and the bearings with three different initial fault locations.

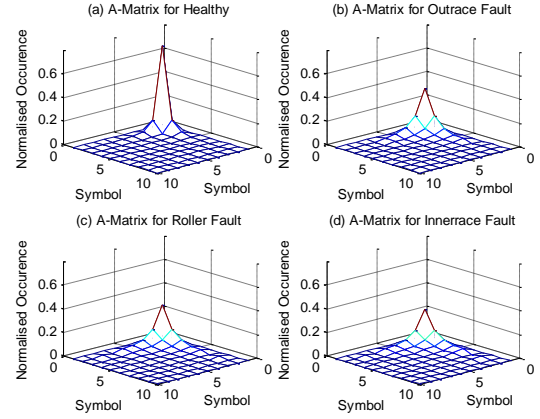


Figure 8. TESPAP A-Matrices for envelope signal of a healthy and three different faulty bearings

Trends in the patterns in A-Matrices shown in Fig. 8 provide very good features to separate fault locations. The healthy pattern has the highest amplitude where the small faulty bearings have much lower amplitudes. The outer race faulty pattern acquired higher amplitude than the other faults; in particular the inner race faulty pattern gained the lowest amplitude.

Therefore, with the introduction of faults the position of the peak changes between the different conditions as displayed in Fig 9. Since the fault causes the peak to change both in position and magnitude such changes will definitely detect incipient faults.

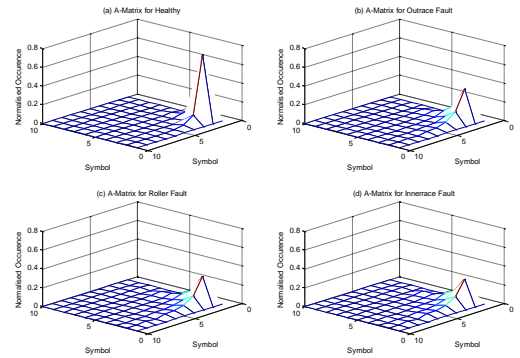


Figure 9. TESPAP A-Matrices for envelope signal of a healthy and three different faulty bearings from different angle

## VII. CONCLUSION

The TESPAP performance in monitoring bearing faults has been assessed with raw data and combined with envelope signals. The results attained from TESPAP and raw signals show that S-Matrices permit fault detection; however, the A-Matrices cannot produce a full diagnosis of the data for classifying fault types. By contrast, the use of TESPAP with envelope analysis both S and A-Matrices allow full defects separation. Hence, this method is most promising and results were obtained show that the combined TESPAP with

envelope approach is more sensitive than either used separately.

#### VIII. REFERENCE

- [1] M.H. Geoge, "TESPAR Paves the Way to Smart Sensor", Sensor Review, MCB University Press., Vol.17, No. 2, 2007.
- [2] V.V. Vu, P.J. Moss, A.N. Edmonds, and R.A. King, "Time Encoded Matrices as Input Data to Artificial Neural Networks for Condition Monitoring Applications". Proceedings of COMADEM '91, Southampton, July 1991.
- [3] G.M. Rodwell and R.A. King, TESPAR/FANN Architectures for low-power, low cost Condition Monitoring Applications. Proceedings of COMADEM '96, Sheffield, July, 1996.
- [4] R.A. King and W. Gosling, Electronic Letters, Vol. 14, pp.456-457, 1978
- [5] R.A. King and T.C Phipps, Shannon, TESPAR and Approximation Strategies, ICSPAT 98, Vol. 18, pp 445-453, Great Britain 1999.
- [6] J.C.R. Licklider, I. Pollack, Effects of Differentiation, Integration and Infinite Peak Clipping upon the Intelligibility of Speech, Journal of the Acoustical Society of America, Vol. 20, no. 1, pp42-51, Jan 1948.
- [7] E.C. Titchmarsh, The Zeros of Certain Integral Functions, Proc. Progress. Math.Soc., Vol. 25, pp. 283-302.
- [8] M.T. Hagan, H.B. Demuth and M. Beale, Neural Network Design, International Thomson Publishing, 1995.
- [9] S. Yang, M.J. Er, and Y. Gao, A High Performance Neural-Network-Based Speech Recognition System, Proceeding of International Joint Conference on Neural Networks, Vol 2, 2001, pp1527.
- [10] L.R. Rabiner and B-H. Juang, Fundamentals of speech recognition. Englewood Cliffs, N.J.: PTR Prentice Hall, 1993.

# Reviewing DSTATCOM for Smart Distribution Grid Applications in Solving Power Quality Problems

Bala Boyi Bukata and Yun Li

School of Engineering,  
University of Glasgow,  
Rankine Building,

Oakfield Avenue, Glasgow G12 8LY, UK.

[b.boyi-bukata.1@research.gla.ac.uk](mailto:b.boyi-bukata.1@research.gla.ac.uk) and [yun.li@glasgow.ac.uk](mailto:yun.li@glasgow.ac.uk)

**Abstract-** Evolution of the future smart distribution grid has made power quality (PQ) resolution a key concern to stakeholders in the power supply sector. The distribution static compensator (DSTATCOM) is one of the custom power devices recently applied to solve this problem. However, manual techniques based on computer-aided-design (CAD) simulator used in designing the DSTATCOM controllers cannot guarantee optimum solutions because of the added integrate-able nonlinearities of the future distribution grid. Therefore, its functionality can be improved through optimum designs by allowing flexible range of working voltages at the point of common coupling (PCC) to meet customer's load demands. This paper focuses on: reviewing DSTATCOM control strategies for solving PQ problems, and its challenges, possible solutions through evolutionary algorithms (EA) based computer-automated-design (CAutoD) of fuzzy logic control, future direction as well as the state-of-the-arts.

**Keywords-** CAutoD, Custom Power, DSTATCOM, Evolutionary Computation, Fuzzy Logic, Power Quality.

## I. INTRODUCTION

Healthy operations in the power distribution system with perfect balance between generated capacity and its demand are a function of optimum power quality (PQ). The inadvertent waveform distortions due to added network complexity defining power quality must have to be restricted through new technologies. Besides, the topic of quality power has practically become a matter of negotiation between suppliers and consumers of electric energy based upon pre-defined standards of cost-benefit analysis [1]. It has been reported that, "poor PQ costs the EU industry and commerce about €10 billion per annum". In the bid to solve the problem, the distribution static compensator (DSTATCOM) based voltage source converter, has recently been actively applied [2]-[6].

Although the static var compensator (SVC) has been applied for PQ improvement since four decades and is

still seen at the transmission corridor for this purpose, the DSTATCOM known for its size advantage, cost effectiveness, high bandwidth and load balancing capabilities, has swiftly taken over this function from the SVC at the distribution level [7]. But the contention whether the DSTATCOM could survive additive pressure due to variability of the renewable energy resources (RES) characterising smart distribution grid also remains. This contention can however be ameliorated by optimizing the DSTATCOM controller through computational intelligence. In this work, a novel learning fuzzy automatic controller (LFAC) is evolved using the fit3pak engine resident in Matlab. While there are various fuzzy control applications of DSTATCOM, yet tuning remains manually based on computer-aided-design (CAD) simulation which takes long time to accomplish [8]-[11]. Additionally, the adaptive feedback loops that come with them do not augment learning procedures necessary for optimisation [12]. In order to achieve design automation, this paper extends the existing CAD simulator to evolutionary algorithm based computer-automated-design (CAutoD) simulation with an augmented learning cycle [13], [50].

Traditional open-loop DSTATCOM controllers built upon pulse width modulation (PWM) and frequency switching techniques have been around since the 1990s for PQ improvement [2], [14], [47], [48]. But, responses from these kinds of controllers do not necessarily reflect desired setpoint following, and can lead to uncontrollable conditions of under/over-voltages. Moreover, the absence of the feedback loop creates an enabling environment for PQ warms to further invade customer's equipment. In order to remove this menace, a comprehensive study in classical closed-loop representation of DSTATCOM controllers for first and higher order models started to emerge in about the same period in mid '90s [15], [51]-[54]. The subsequent decade was littered with literature embodying some of the latest strategies based on modern control and artificial intelligent techniques such

---

This work is fully sponsored by: Petroleum Technology Development Fund (PTDF), Plot 672, Port Harcourt Crescent, Off Gimbiya Street, Off Ahmadu Bello Way, Area 11, Garki, Abuja, Nigeria.



as the proportional-integral-derivative (PID) and fuzzy logic control respectively [16].

The search for PQ solutions has remained vigorous with inclusive reviews on assorted DSTATCOM control techniques listed below. Custom power devices used for enhancing PQ are reviewed with particular reference to some conventional control strategies of the time [15]. The authors in [16] offered a generic survey of artificial intelligence and advanced mathematical tools applied to PQ. Another dedicated review of evolutionary computation in generic electric power systems is also covered by [17]. Recent fuzzy set theory applications in power systems have been detailed in [18]. Power Systems applications of neural networks have as well been reviewed and presented in [19]. Optimisation tools applied to generic power systems control include a comprehensive survey on multi-objective evolutionary optimisation [20] and a survey of particle swarm optimisation [21].

In this paper, the related technologies have been reviewed and concluded for the distribution static compensator (DSTATCOM) of smart distribution grid applications in solving power quality problems. The modern and artificial intelligent DSTATCOM controllers have been introduced such as PID, fuzzy logic, neural networks, and model predictive control (MPC) with the challenges facing them. It can provide a beacon for researchers and design engineers in grasping the state-of-the-arts at a glance.

## II. DSTATCOM CONTROL STRATEGIES

Solutions to some PQ problems such as interruptions, transients and undervoltage or overvoltage may be simply attained through conventional use of capacitor banks, recloser circuit breakers or surge arresters at specific locations. However, for the steady state operation of the network which faces the threats of major problems like sags, flicker, imbalance and harmonic distortions, it is crucial to stack the DSTATCOM with a superfluous structure that would account for the nonlinear dynamics arising from the customer's loads. The following sections present an overview of the control strategies applied in DSTATCOM for improving PQ.

### A. PID Controllers

The PID controller is a three term scheme whose design is coordinated through the action of each of its terms. Ease of implementation and user friendly features makes the PID popular in industrial applications since Elmer's ship autopilot in 1911 [22]. On the other hand, as a linear controller, the PID control performance tends to deteriorate when subjected to a nonlinear process with constantly changing operating conditions and model parameters. However, the current literature has reported the only PID application to reinforcement learning scheme for

studying the control performance of the DSTATCOM [23].

### 1) The PI Controller

The PI controller is often implemented in the series form to balance actuator saturation and remove the need for incorporating a separate anti windup mechanism. The major drawback however is that the PI combination increases the overall system gain, and saturates the integral term as well as introduces low frequency oscillations, which together forces the system out of stability [22]. But this can be settled by passing the excess integral action via a negative feedback. Conversely, the PI controller has the advantage of excellent setpoint following and disturbance rejection at appropriate gain crossover frequency. This advantage coupled with ease of implementation makes it particularly attractive in the DSTATCOM control applications to improve PQ as seen in [24]-[37].

### 2) The PD Controller

The popular view held that, "the derivative term of the PD amalgam destabilises control system", makes it very unattractive and neglected in most PID control applications. The derivative term is perceived to reduce the high frequency roll-off even in the presence of low level noise signal. However, remedies that would make it more attractive exist [22]. So far, there has been no application of the PD scheme to DSTATCOM found in the literature.

### B. Fuzzy logic control

Fuzzy logic control is an aspect of computational intelligence which was first introduced as fuzzy set theory by Lotfi Zadeh in 1965. It was later applied by Mamdani to motor control in 1974. Since then fuzzy control has found significant application in various engineering capacity including the power industry. A bibliography on applications of fuzzy set theory in power systems from 1994 ~ 2001 is covered in [18]. The ability to deal with complex and nonlinear systems gives it an edge over loads of conventional methods. Various regimes of fuzzy-PI, fuzzy-PD and neuro-fuzzy controllers have been successfully developed for the DSTATCOM to solve PQ problems [8], [9], [10], [11] and [38]. The major drawback however with these kinds of controllers is based on manual tuning, except for the neuro-fuzzy breed which has a self-tuning mechanism.

### C. Artificial Neural Networks

Neural networks (NN) are intelligent nonlinear emulators based on the popular black box concept of input-output relationships. They are equipped with training algorithms which enhances their learning capacity to perform as closely as possible the set example through fixed 'weights' and 'bias' terms. They have particularly been used around the world in power systems for load forecasting function. Despite this wide

applicability, NN exhibits a large permutation of training parameters which causes snag during feedforward time-series forecasting that is not conducive for generalization. This criterion increases the sensitivity of the network to a number of choices e.g. size of training sets as detailed in [17]. Nevertheless, applications particular to power quality control using the DSTATCOM are reported [39]-[45]. It is imperative to reiterate that the controllers under discuss are only good for small perturbations, advanced control techniques such as the model predictive control (MPC) discussed below offers ideal responses to large perturbations and are reserved for the future direction.

#### D. Model Predictive Control

Predictive control was better known in petrochemical industry applications. It is now gaining fast ground into other sectors as an advanced control technique. Its unique features of low control update rates, handling actuator constraints and the ability to deal with complex control systems makes it suitable for on-line applications. In PQ prediction, the current DSTATCOM outputs are subjected to some reference signal values based on past inputs in order to calculate the future outputs. This way, any waveform anomaly would have been anticipated and corrected before any disturbance occurs to the customer's device. However, the MPC is yet to be applied, but a harmonic prediction application in DSTATCOM has been reported [46].

#### E. Optimisation Tools

Application of optimisation techniques to the sectors of the power system, especially the area of operation and planning, which is often a large scale problem, has been in existence for more than three decades. A detailed survey on particle swarm optimisation [21] and evolutionary computation applications in general power system are given [21] and [17], respectively. However, their particular application to solve PQ problem in DSTATCOM is still limited. Table 1 presents the pattern of control solutions applied to various PQ problems through the DSTATCOM from 2001 ~ 2011.

TABLE I. CONTROL SCHEME APPLICABLE TO PQ PROBLEMS

Controller \ PQ Problem	PID	PI	PD	FPI	FPD	NN	MPC
Transients	✓	✓	x	✓	x	x	x
Interruptions	x	✓	x	x	x	x	x
Sags/Swells	x	✓	x	✓	x	✓	x
Over/Under Voltage	x	✓	x	✓	x	x	x
Harmonic Distortion	x	✓	x	x	x	✓	x
Flicker	x	✓	x	x	x	✓	x

### III. DSTATCOM CONTROL CHALLENGES

As earlier, most of the techniques applied to curb PQ problems using the DSTATCOM are designed based

on trial-and-error tuning. The methods however do not provide the control unification required for a single controller diversification to offer unrestricted multiple solutions [13]. Hence, the following challenges need to be subdued in order to face the projected complexities of the future distribution grid.

#### A. Intrinsic harmonic generation

One of the serious challenges lies in the self harmonic contribution of the DSTATCOM. Modula design approaches have been successful at the expense of space and cost. Similarly, the use of H-bridge and multilevel converters has managed to reduce the effect of total harmonic distortion and improve PQ control. Moreover, they are mostly implemented off-line with committed open-loop simulation studies based solely on linear models. Serious consideration of nonlinear model simulations must be made to reflect the real time characteristics of the distribution system. The controller to be realised on this format would substantially reduce the total harmonic distortion (THD) content below the allowed limit of  $\pm 5\%$ .

#### B. External harmonic generation

The dynamics of the customer's nonlinear loads constitute one source of external harmonics while the generation systems and neighbouring operational equipment such as protective relays and circuit breakers constitute another. These components continuously keep shifting the operating point of the system and at the same time impose parameter changes within the model. However, the conventional controllers, majorly designed either on linear modelling or manual tuning would not be able to reject all the disturbances at once. This trend would lead to steady state malfunctioning of the network. To reverse the challenge, the evolutionary computation based tuning of DSTATCOM controller will have to be considered.

#### C. Optimisation of the switching angle

One most important challenge in DSTATCOM control is the optimisation of its input signal. The effectiveness of the applied control scheme is a measure of its ability to control desired variable with a minimum control effort. This is hardly achievable especially in an open-loop linear model. The situation even gets worse with a nonlinear model which reduces the magnitude of the control variable due to switching nonlinearities, requiring larger control effort in order to produce the desired output. Therefore, to properly account for these nonlinearities and overcome this challenge, the DSTATCOM design must include a learning loop in addition to the existing feedback and the adaptive feedback loops usually found in its current designs. It is good to note that, the lower the control effort exerted on the switching pattern of the converter IGBT switches, the closer is the setpoint following and the lesser the THD on the output waveform. This is an important index for monitoring PQ problem. Arguably, a learning DSTATCOM would remove the need for

costly multilevel implementations occasioned by the use of PWM and frequency switching techniques.

#### D. Rating of the DSTATCOM

For economic reasons, DSTATCOM rating should be based on voltage deviations at the point of common coupling (PCC) rather than on the magnitude of the load. Equipment manufacturers often create a mismatch between equipment rating and the actual plant rating, so that compensation devices (i.e. DSTATCOM) would have to do extra amount of work to perform some mitigation functions. Consequently, control system designers can reverse the challenge through innovative control designs to improve the ride-through capabilities of the DSTATCOM.

#### IV. EA BASED CAUTOD POSSIBLE SOLUTIONS

The authors proposed possible solutions to the aforementioned DSTATCOM problems based on computer-automated-design (CAutoD) simulator interface. The CAutoD package resident in ft3pak is a nondeterministic optimisation method using evolutionary computation in a given search space. For this work, the DSTATCOM is loaded in the feedback with a simplified three-ruled fuzzy-proportional-derivative (FPD) controller as an adaptive mechanism. A learning mechanism is also incorporated in the system feedback through genetic algorithm (GA). This provides automatic tuning to the parameters of the adaptive mechanism to form the structure called 'learning fuzzy automatic controller (LFAC)' in this project. With this setup, three separate feedbacks are obvious and necessary for realising automatic DSTATCOM control system also called Smart DSTATCOM.

#### V. FUTURE DIRECTION

Prediction Control of the DSTATCOM is being considered as the future direction to this study. Evolution of a simplex predictive control is performed to investigate the characteristics of the DSTATCOM in a wide range of scenarios subject to load disturbances, model uncertainties and violation of hard and soft constraints. The performances are then measured and compared with the preceding techniques.

#### VI. STATE-OF-THE-ARTS

From what we have gathered in this literature review, it is quite obvious that the PI control scheme, with its fuzzy PI variant are the most favoured applications. Despite its disadvantages stated above, the PI is the most widely used strategy to solve PQ problems in the DSTATCOM. This may be explained by the following reasons:

- Ease of implementation.
- Straight forward to use and tune even by control novice.
- Provide fast and accurate results for a lot of PQ control situations.

- Cost effective, available even at over-the-counter stalls.
- Suitable for most linear control applications.
- Effective as ac voltage regulator, current regulator as well as dc voltage controller

#### VII. CONCLUSION

This paper presents a comprehensive review on contemporary control strategies applied to the DSTATCOM to solve PQ problems from 2001 ~ 2011. The challenges facing the schemes and the ways to remedy them have also been highlighted. Possible solutions using GA based CAutoD interface is clearly represented with a rapid follow up of the future direction in the form of evolutionary simplex design of DSTATCOM. State-of-the-arts application has been identified as the PI controller which is thoroughly applicable to both internal and external control loops for ac voltage control, current regulation and dc voltage control. However, the PI controller is reported to collapse in the presence system uncertainties and model parameter changes which warrants the search for new control technologies as the power system evolves into a smart grid.

#### REFERENCES

- [1] D. Chapman; Copper Development Association, "Power Quality and Utilisation Guide Section 5 - Introduction" November 2001, Available From <http://www.copperinfo.co.uk/powerquality/downloads/pqug/51-voltage-dips.pdf>.
- [2] P.R. Sanchez, E. Acha, J.E.O. Calderon, V. Feliu, and A.G. Cerrada, "A Versatile Control Scheme for a Dynamic Voltage Restorer for Power-Quality Improvement," IEEE Transactions on Power Delivery, vol. 24, no. 1, January 2009.
- [3] S.V. Ravi Kumar, "Simulation of D-STATCOM and DVR in Power Systems," ARPN Journal of Engineering and Applied Sciences, vol. 2, no. 3, 2007.
- [4] T.J. Miller, "Reactive Power Control in Electric Systems," John Wiley and Sons, 1982.
- [5] R. Grunbaum, B. Halvarsson, and A. Wilk-wilczynski, "FACTS and HVDC Light for Power System Interconnections," *Power Delivery Conference, Madrid, Spain, September 1999*.
- [6] A. Baghini, Ed., "Handbook of Power Quality," John Wiley & Sons Ltd, 2008.
- [7] H. Masdi, N. Mariun, S. M. Bashi, A. Mohamed, S. Yusuf, "Construction of a Prototype D-Statcom for Voltage Sag Mitigation", European Journal of Scientific Research, Vol. 30 No. 1, 2009, pp.112-127.
- [8] V. Bano, A. Ramirez, M. Juan, "DStatCom regulation by a fuzzy segmented PI controller", Electric Power Systems Research, vol. 80, no. 6, pp 707-715, June 2010.
- [9] Zhu, H. Qun-Feng, Lei; Hu, Zhan-Bin; J. Tang, "The fuzzy PI control for the DSTATCOM based

- on the balance of instantaneous power”, *Emerging Intelligent Computing Technology and Applications - Proceedings 5th International Conference on Intelligent Computing, ICIC v 5754 LNCS*, pp 794-803, 2009,
- [10] Y. Xu, L. Yang, C. Ma, Z. Gong, H. Pu, Z. Zhang, “Study on sliding mode control with RBF network for DSTATCOM”, *2010 International Conference on E-Product E-Service and E-Entertainment (ICEEE) 2010*.
  - [11] R. Coteli, B. Dandil, A. Fikret, “Fuzzy-PI current controlled D-STATCOM”, *University Journal of Science*, v 24, n 1, p 91-99, 2011.
  - [12] J. Sklansky, “Learning systems for automatic Control”, *IEEE Transactions on Automatic Control*, vol. AC-11, NO. 1, January 1966, pp. 6-18.
  - [13] Y. Li, K.H. Ang, G.C.Y. Chong, W. Feng, K.C. Tan, and H. Kashiwagi, “Evolutionary Search and Optimization Enabled Computer Automated Control System Design,” *International Journal of Automation and Computing*, 1(1), pp. 77-78, Oct 2004.
  - [14] N. M. Ndubuka, “Power Quality Enhancement of Nigerian Distribution Systems by Use of Distribution Static Compensator (D-STATCOM)”, *International Journal of Electrical and Power Engineering* 5 (1): pp. 8-12, 2011.
  - [15] Y. Pal, A. Swarup and B. Singh, “A Review of Compensating Type Custom Power Devices for Power Quality Improvement”, *Joint International Conference on Power System Technology POWERCON and IEEE Power India Conference, POWERCON 2008*.
  - [16] W. R. Anis Ibrahim, M. M. Morcos, “Artificial Intelligence and Advanced Mathematical Tools for Power Quality Applications: A Survey”, *IEEE Transactions on Power Delivery*, Vol. 17, NO. 2, April, 2002.
  - [17] A. P. Alvas da Silva, “Applications of Evolutionary computation in Electric Power System”, *Proceedings of the 2002 Congress on Evolutionary Computation, CEC'02, Vol. 2, May 2002*, pp. 1057-1062.
  - [18] R. C. Bansal, “Bibliography on the Fuzzy Set Theory Applications in Power Systems (1994-2001)”, *IEEE Transactions on Power Systems*, Vol. 18, No. 4, November 2003.
  - [19] M. Tarafdar Haque and A. M. Kashtiban, “Application of Neural Networks in Power Systems: A Review”, *Proceedings of World Academy of Science, Engineering and Technology, Volume 6, June 2005*, pp. 53-57.
  - [20] N. M. Pindoriya, S. N. Singh, and K. Y. Lee, “A Comprehensive Survey on Multi-objective Evolutionary Optimisation in Power System Applications”, *IEEE Power and Energy Society General Meeting*, July 2010, pp. 1-8.
  - [21] M. R. AlRashidi and M. E. El-Hawary, “A Survey of Particle Swarm Optimisation Applications in Electric Power Systems”, *IEEE Transactions on Evolutionary Computation*, Vol. 13, No. 4, August 2009, pp. 913-918.
  - [22] Y. Li, K. H. Ng and G. C. Y. Chong, “PID Control System Analysis and Design”, *IEEE Control Systems Magazine* 26(1): 2004, pp. 32-41.
  - [23] M. Xing-ping, W. Hui, Z. Liang, Z. Hong, “Controlling study of D-STATCOM based on reinforcement learning adaptive PID,” *Automation and Logistics, 2009. ICAL '09. IEEE International Conference on*, vol., no., pp.1208-1211, 5-7 Aug. 2009.
  - [24] W. Dai, “DSTATCOM inverse system PI controller”, *Advanced Measurement and Test*, v 439-440, pp 372-377, 2010.
  - [25] S. Harish; M. K. Mahesh, “Fuzzy logic based supervision of DC link PI control in a DSTATCOM”, *Proceedings of the INDICON 2008 IEEE Conference and Exhibition on Control, Communications and Automation*, v 2, pp 453-458, 2008,
  - [26] J. Tang, Z-Y. Deng, Z-H. Yang, “Double closed loop control for distribution static synchronous compensator based on state PI feedback control”, *Journal of Central South University (Science and Technology)*, v 41, n 6, p 2282-2287, December 2010.
  - [27] B. Singh, A. Adya, A. P. Mittal, J. R. P. Gupta, “Modeling, design and analysis of different controllers for DSTATCOM”, *Joint International Conference on Power System Technology POWERCON and IEEE Power India Conference, POWERCON 2008*
  - [28] Fan, S. Ruixiang, Min; F. Shen, C. Tu, “Frequency dividing coordinated control method for DSTATCOM based on the limit values judgment”, *Automation of Electric Power Systems*, v 33, n 4, pp 67-71+91, February 25, 2009
  - [29] N. Farokhnia, S. H. Fathi, H. Toodeji, “Direct nonlinear control for individual DC voltage balancing in cascaded multilevel DSTATCOM”, *International Conference on Electric Power and Energy Conversion Systems, EPECS 2009*
  - [30] R. Majumder, A. Ghosh, G. Ledwich, F. Zare, “Power sharing and stability enhancement of an autonomous microgrid with inertial and non-inertial DGs with DSTATCOM”, *International Conference on Power Systems, ICPS '09*.
  - [31] Kumar, R. Dinesh “Modelling, analysis and performance of a DSTATCOM for unbalanced and non-linear load”, *IEEE/PES Transmission and Distribution Conference and Exhibition: Asia and Pacific*, v 2005, pp 1-6, 2005.
  - [32] C. N. Bhende, M. K. Mishra, H. M. Suryawanshi, “A DSTATCOM modelling, analysis and performance for unbalanced and non-linear load,” *Journal of the Institution of Engineers (India): Electrical Engineering Division*, v 86, n MAR., pp 297-304, March 2006.

- [33] B. Singh, A. Adya, A. P. Mittal, J. R. P. Gupta, "DSTATCOM for power quality improvement in a four-wire electric distribution system", *International Journal of Global Energy Issues*, v 26, n 3-4, pp 401-416, 2006.
- [34] G. Marcelo P. E. Mercado, "Power flow control of microgrid with wind generation using a DSTATCOM-UCES", *Proceedings of the IEEE International Conference on Industrial Technology*, p 955-960, 2010.
- [35] J. Segundo-Ramirez, A. Medina, A. Ghosh, G. Ledwich, "Stability analysis based on bifurcation theory of the DSTATCOM operating in current control mode", *IEEE Transactions on Power Delivery*, v 24, n 3, p 1670-1678, 2009.
- [36] N. Prabhu, M. Janaki, R. Thirumalaivasan, "Control design and performance evaluation of energy source interfaced Dstatcom", *IET-UK International Conference on Information and Communication Technology in Electrical Sciences, ICTES v 2007*, n 2, p 517-522, 2007.
- [37] C. Tang, "Nonlinear control method of DSTATCOM", *Dianli Zidonghua Shebei/Electric Power Automation Equipment*, v 31, n 3, p 18-22+23, March 2011.
- [38] E. Deniz, S. Tuncer, R. Coteli, F. Ata, B. Dandil, M. T. Gencoglu, "Neuro-fuzzy current controller for three-level cascade inverter based D-STATCOM" *Proceedings of the Universities Power Engineering Conference, 45th International Universities' Power Engineering Conference, UPEC 2010*.
- [39] X. P. Yang, W. Yan-Ru; Yan, "A novel control method for DSTATCOM using artificial neural network", *Proceedings of the IPEMC: CES/IEEE 5th International Power Electronics and Motion Control Conference v 3*, pp 1724-1727, 2007.
- [40] B. Singh, J. Solanki, V. Verma, "Neural network based control of DSTATCOM with rating reduction for three-phase four-wire system", *Proceedings of the International Conference on Power Electronics and Drive Systems*, v 2, pp 920-925, 2005.
- [41] B. Singh, A. Adya, A. P. Mittal, J. R. P. Gupta, "Neural network based DSTATCOM controller for three-phase, three-wire system", *International Conference on Power Electronics, Drives and Energy Systems, PEDES '06*
- [42] S. Srivastava, K.S. Preeti, D. J. Sharma, M. Gupta, M. "Online estimation and control of voltage flicker using neural network", *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2009.
- [43] B. Singh, P. Jayaprakash, S. Kumar, D. P. Kothari, "Implementation of neural network controlled three-leg VSC and a transformer as three-phase four-wire DSTATCOM", *IAS Annual Meeting (IEEE Industry Applications Society)*, *IEEE Industry Applications Society Annual Meeting*, 2009
- [44] M. S. Karami, A. Heidar, A. G. Tapeh, S. Bandari, "Learning techniques to train neural networks as a state selector in direct power control of DSTATCOM for voltage flicker mitigation", *Proceedings of the International Conference on Information Technology: New Generations (ITNG)*, p 967-974, 2008.
- [45] X-P. Yang, Y-R. Zhong, Y. Wang, "A novel control method for DSTATCOM using artificial neural network", *Conference Proceedings of the CES/IEEE 5th International Power Electronics and Motion Control Conference*, v 3, p 1724-1727, 2007.
- [46] F. Ruixiang, L. An, S. Min, S. Zhikang, "Harmonic prediction control for time delay canceling and its realization in distribution static synchronous compensator", *Proceedings of the WRI Global Congress on Intelligent Systems (GCIS)*, v 3, pp 56-60, 2009.
- [47] V. George, M. K Mishra, "Design and analysis of user-defined constant switching frequency current-control-based four-leg DSTATCOM", *IEEE Transactions on Power Electronics*, v 24, n 9, p 2148-2158, 2009.
- [48] R. Fan, M. Sun, F. Shen, C. Tu, "Frequency dividing coordinated control method for DSTATCOM based on the limit values judgment", *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, v 33, n 4, p 67-71+91, February 25, 2009.
- [49] T. Manmek, C. P. Mudannayake, C. Grantham, "Voltage dip detection based on an efficient least squares algorithm for D-STATCOM application", *Proceedings of the CES/IEEE 5th International Power Electronics and Motion Control Conference (IPEMC)*, v 2, pp 1207-1212, 2007.
- [50] B. B. Bukata, "Evolving Optimum Fuzzy Logic Controllers for Power Distribution Systems", A 2<sup>nd</sup> Year Report Submitted to the School of Engineering, University of Glasgow, May, 2010.
- [51] A. V. Jouanne and B. Banerjee, "Assessment of voltage unbalance," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 782-790, Oct. 2001.
- [52] P. Selvan and R. Anita, "Transient Enhancement of Real Time System Using STATCOM", *European Journal of Scientific Research*, Vol. 52, No.3, pp.359-365, 2011
- [53] S. Goyal, A. Ghosh, G. Ledwich, "A hybrid discontinuous voltage controller for DSTATCOM applications", *IEEE Power and Energy Society 2008 General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century, PES*
- [54] S. Y. Jung, T. H. Kim, M. U. Lee, I. S. Moon, W. H. Kwon, "Modeling and control of DSTATCOM for voltage sag", *Proceedings of the IASTED International Conference on Modelling, Simulation and Optimisation*, pp 108-113, 2003



# *Analysis of interrelationships between Suppliers configuration and performance within the supply network: A simulation approach*

Maria Aina\*, Dr. Yang Dai\*, Professor Dobrila Petrovic  
Department of Engineering and Knowledge Management, Coventry University  
Coventry, UK

\* Contact Email: [ainam@coventry.ac.uk](mailto:ainam@coventry.ac.uk); [y.dai@coventry.ac.uk](mailto:y.dai@coventry.ac.uk)

**Abstract** — The objective of this work is to examine the strategic types of organisations suppliers. Moreover, it is to investigate the interrelationship between various ratios of suppliers' strategic types and the effect on the performance of the hub company in a supply network. Also, to suggest the configuration ratio that produces enhanced performance. The supply network considered in this study is the Multiple Driven supply network of Dai and Zhang's [4] supply network. Three indicating variables were used for grouping the suppliers' strategic types based on Miles and Snow typologies [11]; Defender, Prospector and Analyser. This is a case study of 30 printing organisations suppliers of a Multiple Driven Supply Network with three different Miles and Snow typologies [11]. The choice of measuring of the organisation's performance is the quantity produced and price. Questionnaires, Semi- structured interviews and the Discrete Simulation Software were adopted for analysing the hypotheses and to suggest the improvement of the organisations performance. The first experimental work is the 'AS IS' scenario which is the real situation of the hub organisation and its suppliers. This is followed by the 'WHAT IF' scenario which are; All Defenders, All Analyser, All Prospectors, Ratio (1:1:1) and (3:2:1) of (Analyser: Defender:Prospectors). Simulation provides the performances of the scenarios both in terms of profit and the quantity produced. The suppliers typological ratio that gives enhanced performance in terms of quantity produced is (3:2:1) with respect to (Analyser: Defender: Prospector). However, the profit is the greatest when the suppliers are all Analysers.

**Keywords**-Multiple Driven Supply Network; Miles and Snow typologies; network configuration; business simulation

## I. INTRODUCTION

In the last decade, competition in modern business environments has shifted increasingly from 'Organisation vs. Organisation' towards 'Supply chain vs. Supply chain'. This has made supply chain and supply chain management become more important. Supply chains consist of independent, partially discrete, value-adding units that work together to transform raw materials into finished products[2]. Supply chain management involves the coordination of manufacturers, suppliers, distributors

and retailers to provide customers with the value added products or services[18]. Failure to manage supply chains presents serious negative consequences. For example, this resulted in Motorola's loss of crucial early sales of camera phones in the year 2003. In 2001; it resulted in Cisco to writing off \$2.25 billion in inventory [10]. However, firms such as Wal-Mart, Zara, Toyota, and Dell have effectively used supply chain management as a tool to improve key outcomes, performances and to gain competitive advantages over their peers. Activities relevant to supply chain management include integrating supply chain into the system, designing the supply chain, inventory control, management of information flow, customer service, integrating planning and control system, financial and physical flows, and supply chain configuration [9].

This paper centres on the configuration aspect and aims to suggest the efficient configuration that produces improved performance in a specific supply network [18]. It uses a Multiple Driven Supply Network which focuses on the first suppliers, the core organisation and a manufacturing process. The hub company of this supply network is an Analyser which is peculiar because it shares the characteristics of a Prospector and Defender and has received little attention in literature. The ultimate goal of this study is to assist companies to form appropriate supply chains by choosing the right suppliers with the most added-value to customer order fulfillment. Various configurations of the supply chain based on the ratio of supplies typologies are analysed using SIMUL 8. A case study of a printing firm that produces label is used to illustrate different supplier's typological ratio and their effects on the core organisation's performance.

## II. BACKGROUND

The increased interest in supply chain has led to the development of various models and tools that aims to support the design, configuration and analysis of supply chain networks [10]. [17] developed an approach based on mathematical programming to optimise supply chains that

focused on the product structure. [16] focused on the configuration of supply chain relative to various degrees of product customisation. A three-layered decision support system was designed by [5] for supply chain configuration, which assists in the selection and evaluation of supply chain entities. [3] developed a decision support modeling methodology, which centres on information about inventory, lead time and item design and analysed their impact on supply chain performance. These works on configuration ignored the choice of suppliers within the supply network except for [13] that proposed an Automated Supply Chain Configurer (ASCC) framework applicable to a company to select its immediate suppliers. However, it pays no attention to the effect of the choice of suppliers on the organisations performance. An important aspect of supply chains is to design and configure them in such a way that it reaches optimal performance. The choice of suppliers within a supply chain eventually leads to the overall system performance. The most important task in the supply chain configuration is to allocate resources and select suppliers [7]. This work models around the typology of the suppliers, considering the effect of their cost and time of delivering raw materials on the hub organisations performance (number of finished labels and profit).

This study imbibes one of the three supply network models generated by Dai and Zhang [4]; Cost Saver, Adapter and Multiple Driven. The concept is based on 'Hub and spokes' where the core organisation is said to be the 'Hub' and the suppliers and customers are the 'Spokes'. The first tier suppliers and customer orbits around the hub company which makes this network easier to manage. Also, the organisations within the supply network are grouped based on the Miles and Snow's typology.[11] considered typologies labelled 'Defenders' which emphasises on reduced cost. They are functionally organised and focuses on efficiency (doing things right) whilst avoiding unnecessary risk. 'Prospectors' invest in high-tech products which are highly priced, they are externally orientated, and centres on effectiveness (doing the right things), and 'Analysers' is a hybrid of the two former typologies. They strive to balance effectiveness and efficiency, whilst delivering the highest quality products at reasonable prices. This unique concept that groups organisations with their strategic typologies still lies latent in literature. This study, however, intends to satisfy this omission pointed out by [14], the need to identify typology and examine their relations with dependent variables.

### III. THEORY AND HYPOTHESES

Aina, Dai, and Petrovic (2010)'s study reveals that the ratio of Defender: Prospector: Analyser cannot be equal in a multiple driven supply network i.e. (1:1:1). Based on this, two hypotheses has been proposed;

**Hypothesis 1:** If the hub company in a Multiple Driven Supply Network is an Analyser, and if majority of the

suppliers are Analysers, then higher performance can be achieved compared to other configurations.

**Hypothesis 2:** If a Multiple Driven Supply Network has much less Prospectors than Defenders, then the suppliers will have reduced performance.

### III. METHODOLOGY

Since the choice of suppliers within the supply chain has an effect on customers, it is essential for an organisation to identify the one that leads to the best added-value. Simulation has been proven as a promising analysis tool to assist decision makers in various problems [19]. In this research, a new Simul8 model is developed to simulate supply chain configuration and its activities and allows various *what-if* analyses to be carried out. It allows the comparison of different operational alternatives without disrupting the real system.

This is a case study of a printing organisation that produces labels. The questionnaires were distributed to 50 suppliers, but only 30 responded. The questionnaire was developed to deduce the strategic typology of each supplier; the hub organisation was interviewed to know the efficiency and the cost of goods delivered by their suppliers. Refer to (Fig. 1), for the flow of events within the printing firm. This diagram is represented on the simulation software to run the processes of the organisation and its suppliers.

### IV. RUNNING EXPERIMENTALWORK

The 'AS IS' simulation of the network (the real state printing press and its suppliers) was compared to the performances of the five 'What if' Scenarios below. These scenarios are considered to help determine the ratio that enhances the hub organisations performance of the Multiple Driven Network.

- **Scenario 1;** *What if* 'All Multiple Driven supply network suppliers are Defenders'
- **Scenario 2;** *What if* 'All Multiple Driven supply network suppliers are Analysers'
- **Scenario 3;** *What if* 'All Multiple Driven supply network suppliers are Prospectors'
- **Scenario 4;** *What if* 'The ratios of the strategic types of suppliers of the Multiple driven supply network are (1:1:1) with respect to Analyser: Defender:Prospector
- **Scenario 5;** *What if* 'The ratios of the strategic types of supplier of the Multiple Driven supply network are (3:2:1) with respect to Analyser: Defender: Prospector

### IV. RESULT AND ANALYSIS

A close study on the printing organisation was carried out for 3 months to observe the suppliers effectiveness and

efficiency (the suppliers date and time agreed for the materials to be delivered, time of delivery by the supplier and cost of raw materials). However, most raw materials were delivered to the organisation once every month to produce labels. The result reveals that Defender suppliers have the most precise (as agreed with the hub *Organisation*) delivery time while the Prospector and Analysers supplier organisation are not most times precise. It was also observed in the data collected that, for same products that are supplied by Defender, Prospectors or Analyser suppliers, prices of Prospectors are greater or equal to the prices of Analysers, which are in turn greater than the prices of Defenders.

SIMUL8 contains a wide range of Statistical Distributions which provide a method of simulating variations that occur in timing in any process involving people or organisations. The data for Prospectors show a large variability in delivery times, Defenders comes in at the exact times having fixed distribution in delivery times, while the Analysers show randomness in delivering to the hub organisation. The simulation was set to run for 35 days (7 weeks), for a period of 5 working days per week between the hours of 8am to 5pm. The variables considered are the cost per unit of goods delivered, quantity delivered and the time delivered. However, the performances are measured by the quantity of finished products and the profit made within the set time.

Below are the results of the 'AS IS' and 'WHAT IF' (are assumptions carried out to test the hypotheses) scenarios. The Revenue per unit in Simul 8 is described as the cost of each finished products. The tables below show the amount of finished products that was produced in 7 weeks, the number of labels produced, the total cost of the raw materials, the total revenue (Number of finished products \* Revenue per unit), the Profit is calculated as (Total Revenue - Total Cost) and the profit per unit.

In simulating the 'AS IS' scenario, the time each suppliers delivered the raw materials and the cost per unit were input into the simulation model. The results are shown in (Table I) below;

TABLE I. THE 'AS IS' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
821	10,425.75	41,050.00	30,624.25	37.30

#### THE 'WHAT IF' RESULTS

**Scenario 1:** The data for Analysers and Prospectors suppliers in the Multiple Driven supply network were replaced by Defenders data. The result is shown in (Table II) below;

TABLE II. THE 'ALL DEFENDERS' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
802	23,799.06	40,100.00	16,300.94	20.33

**Scenario 2:** The data for Defenders and Prospectors suppliers were substituted with Analysers data and the result is shown in (Table III) below;

TABLE III. THE 'ALL ANALYSERS' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
814	9,735.56	40,700.00	30,964.44	38.04

**Scenario 3:** The data for Defenders and Analysers suppliers were substituted for Prospectors. The result is shown in (Table IV) below;

TABLE IV. THE 'ALL PROSPECTORS' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
796	9,896.56	39,800.00	29,903.44	37.57

**Scenario 4:** The three typologies were ascribed the same number of suppliers i.e. *ratio (1:1:1)* with respect to Analyser: Defender: Prospector in the Multiple Driven supply networks. The result is shown in (Table V) below;

TABLE V. THE 'RATIO 1:1:1' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
795	9600.29	39,167.46	29,567.17	37.19

**Scenario 5:** The result shown in (Table VI) below was obtained when the ratios of the Multiple Driven supply network suppliers are 3:2:1 (Analyser: Defender: Prospector)

TABLE VI. THE 'RATIO 3:2:1' RESULT

Number of labels produced	Total cost (£)	Total Revenue (£)	Profit (£)	Profit /Unit (£)
821	10,425.75	41,050.00	30,624.25	37.30

Fig. II and Fig. III are the graphical representation of the results from each Scenario. This shows the profit and the quantity produced respectively.

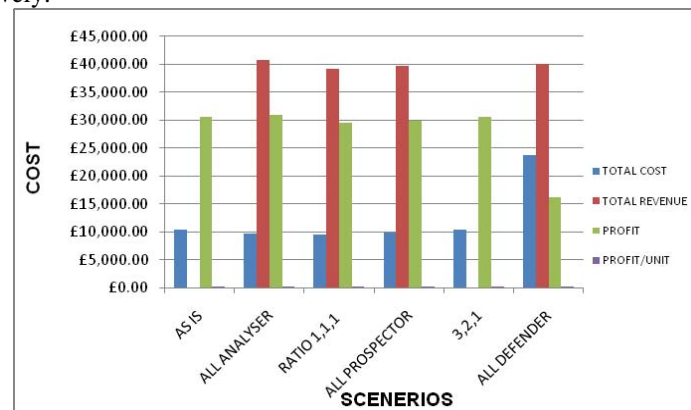


Figure II. Graphical representation of performance (Profit)

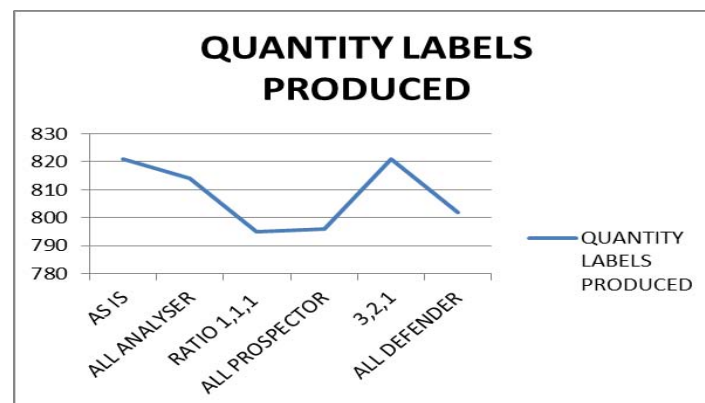


Figure III. Quantity of labels produced

The result for the 'AS IS' scenario is same as for the ratio 3:2:1. This is because the suppliers' configuration for the hub company is ratio of 3:2:1 (Analyser: Defender: Prospector).

**Hypothesis I:** If the Hub company in a Multiple Driven supply network is an Analyser, and if majority of the suppliers are Analysers; then higher performance can be achieved compared to other configurations. The result for Scenario 'All Analyser' shows the highest profit per unit of £38.04 with 814 labels produced. Also, Scenario '3:2:1' (Analyser: Defender: Prospector) produced (821) the highest number of labels within the set period of time and a profit of (£37.30) per unit of labels. These results support hypothesis I.

**Hypothesis II:** A Multiple Driven supply network that has much less Prospectors than Defenders as the suppliers will have reduced performance. As shown in the results above, when the organisations suppliers are 'All Prospectors' the profit per unit of labels was considerably high (£37.57) compared to (£20.33) of All Defenders. However, the number of finished labels after 7 weeks are higher for Defenders (802) while for

Prospectors it is (796). This results indicates that both typologies strength lies in different performances. Therefore, hypothesis II, is accepted. However, the Hub organisation is left with the choice either to allow more Prospectors in order to make more profit, or to allow more Defenders to satisfy customers high demand for labels.

## V. CONCLUSION

The results of this study shows that the performance in terms of the quantity produced (labels) produced was highest when the suppliers configurations are 3:2:1 (Analyser: Defender: Prospector), while the performance in terms of profit was highest when All suppliers were Analysers. However, the lowest number of labels were produced when the ratio of Suppliers are 1:1:1 (Analyser: Defender: Prospector) and the profit was Lowest when All suppliers were Defenders. This paper shows analyses of various ratios and their subsequent performances. It suggests that the Hub organisation should select suppliers to meet its performance (Quantity or Profit) needs. These studies encourage further works into 'The effect of the quality of products delivered by the suppliers, on the quality of the end product within the Multiple Driven Supply Network'.

## REFERENCES

- [1] Aina M., Dai Y., and Petrovic D., 2010 A sustainable supply network model and its configurations: A novel supply network model for the hub and spokes companies, Coventry University, UK
- [2] Bowersox DJ, Closs DJ, Stank TP. 1999. *21st Century Logistics: Making Supply Chain Integration a Reality*. Council of Supply Chain Management Professionals: Oak Brook, IL.
- [3] Blackhurst, J., Wu, T., and O'Grady, P., 2005. PCDM: a decision support modeling methodology for supply chain, product and process design decisions. *Journal of Operations Management*, 23 (3–4), 325–343.
- [4] Dai, Y., and Zhang, Z. (2008) 'The technological preference of the hub company in a supply network' *Proceedings of the 14th International Conference on Automation & Computing, Brunel University*, 14(2-3)
- [5] Dotoli, M., Fanti, M.P., and Meloni, C., 2003. A decision support system for the supply chain configuration. In: *Proceedings of the*
- [6] 2003 IEEE international conference on systems, man and cybernetics, 5–8 October 2003, Washington, USA, 2667–2672.
- [7] Graves, S.C. and Willems, S.P., 2003. Optimizing the supply chain configuration for new products. Working paper. Leaders for Management Program and A.P. Sloan School of Management, MIT.
- [8] Hult, GTM, Ketchen DJ, Slater SF. 2004. Information processing, knowledge development, and strategic supply chain performance.
- [9] Huang, S.H., Uppal, M., and Shi, F., 2002. A product-driven approach to manufacturing supply chain selection. *Supply Chain Management: An International Journal*, 7 (4), 189–199.
- [10] Lee HL. 2004. The triple-a supply chain. *Harvard Business Review* 83: (October): 102–112.
- [11] Miles, R. E. and Snow C.C. (1978), *Organizational Strategy, Structure, and Process*. Palo Alto, CA, USA: Stanford University Press
- [12] Miller, D. and H. Mintzberg (1983). 'The case for configuration'. In G. Morgan (ed.), *Beyond Method: Strategies for Social Research*. Sage, Newbury Park, CA, pp. 57–73.
- [13] Piramuthu, S., 2005. Knowledge-based framework for automated dynamic supply chain configuration. *Production, Manufacturing and Logistics*, 165, 219–230.
- [14] Rumelt, R. P., D. E. Schendel and D. Teece (1994). *Fundamental Issues in Strategy: A Research Agenda*. Harvard Business School Press, Boston, MA.
- [15] Sahin, F. and Robinson, E., 2002. Flow coordination and information sharing in supply chains: review implications and directions for future research. *Decision Sciences*, 33 (4), 505–536.
- [16] Salvador, F., Rungtusanatham, M., and Forza, C., 2004. Supply-chain configurations for mass customization. *Production Planning & Control*, 15 (4), 381–397.
- [17] Yan, H. and Yu, Z., 1998. A strategic model for supply chain design with logical constraints: formulation and solution. Working paper, No. 04/98–9
- [18] Yan, H., Yu, Z.X., and Cheng, T.C.E., 2003. A strategic model for supply chain design with logical constraints: formulation and solution. *Computers & Operations Research*, 30 (1), 2135–2155.
- [19] Yoon Chang and H. Makatsoris, (2001), "Supply Chain Model-ing Using Simulation", *International Journal of Simulation: Systems, Science & Technology*. (Simulation society), Vol 2, No1, p24-30.



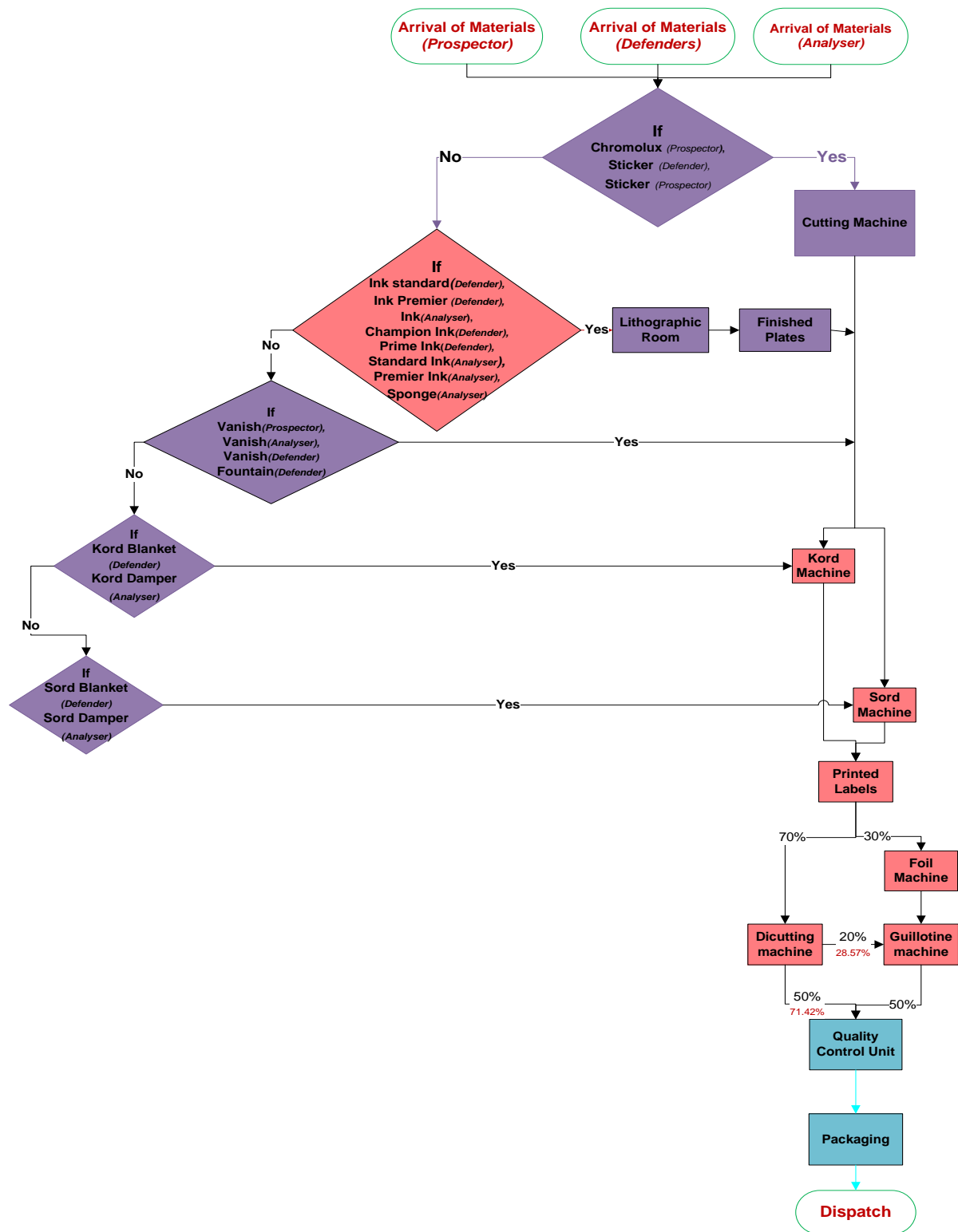


Figure I. Flow chart of the hub organisation processes

# Review Electromagnetic Field Ignition System for Internal Combustion Engine

Lutz-Christoph Schöning\* and Yun Li\*\*

School of Engineering  
University of Glasgow  
Rankine Building

Oakfield Avenue, Glasgow G12 8LT, UK

\* [lschoening.l@research.gla.ac.uk](mailto:lschoening.l@research.gla.ac.uk) and \*\* [yun.li@glasgow.ac.uk](mailto:yun.li@glasgow.ac.uk)

**Abstract** – the Homogeneous Charge Microwave Ignition (HCMI) system combines the advantages of a Spark Ignition- (SI) and Compression Ignition (CI) System. Petrol Internal Combustion Engines (ICEs) uses a spark ignition system, which currently offers only energy efficiency between 25% and 35% [1]. It also produces excessive exhaust emissions. The fact of using the advantages of both common ignition techniques makes a significant impact for the fuel consumption and emission of the combustion process. Through the HCMI system, the fuel inside of the engine cylinder will ignite time simultaneously, which improves the engine efficiency significant. Just four academic groups (Glasgow, West Virginia, Southern California and Pisa) are breaking research on microwave ignition engines. The University of Glasgow is one of the first groups to carry out any computer simulations of a Microwave Ignition (MI) system and the first to use computational intelligence for the design to bring about this revolution.

**Keywords** – *electromagnetic field ignition system, internal combustion engine, microwave Ignition*

## I. INTRODUCTION

The common Petro Engine (Otto Engine) was used for centuries, while the Otto Engine performance improves over the years, the ignition system is nearly the same. The homogeneous air-fuel mixture ignites through a high voltage impulse (mostly 10kV - 25kV) to the spark plug. The ignition occurs between the electrodes (2mm) at just one point. Because of this one point ignition the SI has the disadvantage that not the complete air-fuel mixture ignites. This fact impacts the fuel efficiency and exhaust emission.

From 1978 to 2000 [2] the energy efficient of a SI engine improved fast by around 18% but in the years 1990 to 2000 this process became slower and improved by just 1% of the total value. This statistic data shows that the room for further improvements for a common SI engine is negligible.

To improve the efficiency of one point ignition, researchers investigated multi point ignition [3] [4]. For this technique the common one spark plug is replaced by more spark plugs, typical three. Experiment results shown that the potential is minimal, due to the ignition time synchronization between the separate spark plugs. HCMI is not the first technique which tries to combine the advantage

of SI and CI systems. In 1983 Nojt [5] proposed the idea of Homogeneous Charge Compression Ignition (HCCI) engines. The HCCI technique shows that it can replace the common SI system over the wide range. During the last years, a lot of research occurs in this area [6] [7] [8], the problem of HCCI engines is the control of the ignition event. A HCCI engine has no significant and well defined combustion indication which can be used to control the combustion.

## II. MICROWAVE IGNITION

The MI system was first proposed and patented 1974 by Ward [9]. This patent described that through a microwave transmission line and a microwave emitter, microwave energy is coupled into the engine cylinder. The microwave resonance inside the combustion chamber generates a strong electric field. This field breaks down the injected air-fuel mixture inside the cylinder.

Since this time several patents and research papers have been published and submitted in this area.

All these different tries and ideas can be spitted into two separate groups. One is using the engine cylinder as a resonator as described above [10] [11] [12] [13] [14] ; the other idea is to develop an independent resonator for the air-fuel mixture ignition inside the cylinder [15] [16].

In 1997 DeFreitas [17][18] patented two different ignition apparatus (one shown in Figure 1) for a combustor includes a microwave energy source. The microwave energy will transmitted into the combustion chamber at a resonance frequency.

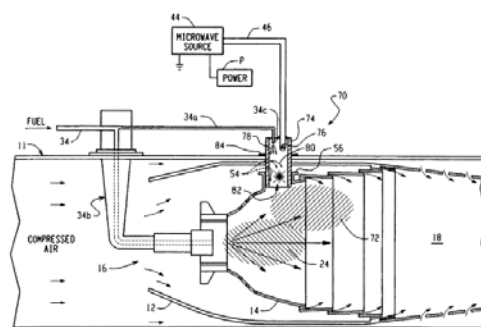


Figure 1 Ignition Apparatus [17]

Through the movement of the piston the natural frequency of the cavity changes; to solve this is the main challenge of this principle, independent of the engine cylinder geometries [10] [11] [12] [14].

Initialized by this problem, researchers at West Virginia University [16] developed a quarter wave coaxial cavity resonator (QWCCR). QWCCR works as an independent resonator; the generated strong electric field will ignite less air-fuel mixture to save fuel and improve the energy efficiency. QWCCR was developed to replace the spark plug without mechanical change on the engine itself. Unfortunately the ignition by using QWCCR occurs just around the centre electrode, which dissipated the advantage of using microwave ignition.

Comparable to QWCCR an RF plasma ignition device for ICE (with a frequency range of 800-1500MHz) was developed by Smith [19] by using a QWCCR in a special cylinder.

2002 Schleupen [11] patented an ignition device for a high frequency ignition.

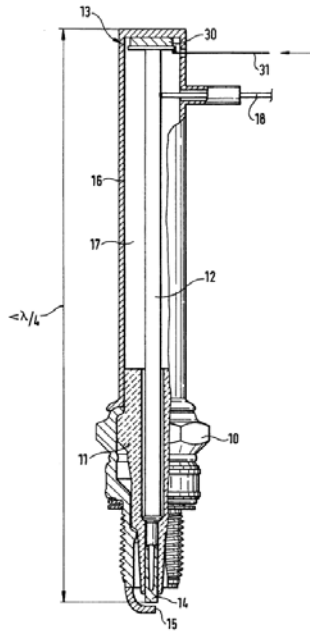


Figure 2 Device for a High Frequency Ignition [20]

Figure 2 shows the principle of the radio frequency ignition. The illustrate plug can just replace the common spark plug without mechanical changes for the engine.

Dana Corporation demonstrated in 2005 a microwave based on plasma ignition technology, AtmoPlas<sup>(TM)</sup>. AtmoPlas combine the micro-wave source feeds with short microwave pulses into the engine though a modified spark plug. The ignition occurs with a plasma temperature of 1200 degree Celsius (no known upper practical level), nevertheless no details information on AtmoPlas has been published.

MWI (Micro Wave Ignition) GmbH published patents [21] [22] for a MI system between 2003 and

2005. In 2006 MWI was planning to build a prototype in cooperation with different automobile manufacturers within two years and within ten years to supply every new automobile with their technology [23].

In 2010 Makita and Ikeda [24] patented an apparatus for ignition or plasma generation. The purpose is to eliminate the need for resonance in the combustion chamber.

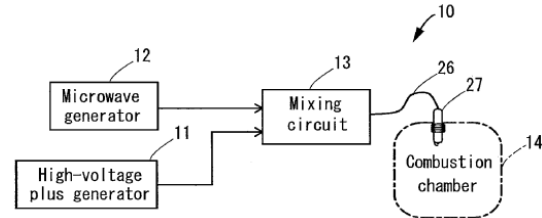


Figure 3 Ignition or Plasma Generation Apparatus [24]

The research in the past shows the feasibility and advantage of MI systems [25]. So far no one has successfully realised a working MI system. Throughout the used volume ignition MI completes the combustion and reduces the fuel consumption and exhaust emissions.

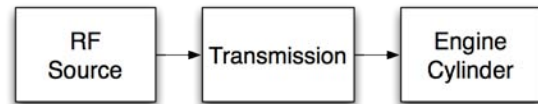
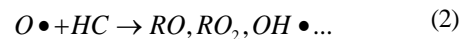
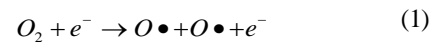


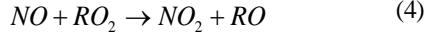
Figure 4 MI Source to Engine

After the control signal (ignition event) the microwave is generated by the microwave source. Through the transmission line (for the engine application a coaxial cable) the microwave will transfer to the engine cylinder. MI works with a multipoint ignition, more specified a volume ignition), also like a Homogeneous Charge Compression Ignition (HCCI) but eliminates the problem to control the ignition event. The control of the ignition is as easy as the control of SI systems.

#### A. Plasma Chemistry

Plasma discharges forms electrons, positively charged ions and free radicals. Air-fuel mixture consists primarily of N<sub>2</sub>, O<sub>2</sub>, NO, NO<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, CO, CO<sub>2</sub>, H<sub>2</sub>O, non-combusted fuel feedstock, and particular matter [26] [27].





R is an organic species and • denotes radical species. Ozone is formed as an intermediary of the reaction described below:



### B. Microwave Fundamentals

The Maxwell equations below are used to describe the EM (electric and magnetic) field inside the combustion chamber [28].

$$\nabla \bullet D = \rho \quad (6)$$

$$\nabla \bullet B = 0 \quad (7)$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad (8)$$

$$\nabla \times H = J + \frac{\partial D}{\partial t} \quad (9)$$

Where  $E$  [V/m] is electric field intensity,  $H$  [A/m] is magnetic field intensity,  $B$  [T] is magnetic flux density,  $D$  [C/m<sup>2</sup>] is electric displacement field,  $J$  [A/m<sup>2</sup>] is current density, and  $\rho$  [C/m<sup>3</sup>] is electric charge density. Different propagation modes can be used to supply an EM field inside the combustion chamber. Depending on the application the modes supply just an electric or magnetic field, both or none of them.

The resonance frequency is an important factor in a MI system. Through the moving of the piston, changing the geometries of the resonance chamber, the resonance frequency will change all the time. At this frequency the microwave resonates and produces a strong electric field inside the cylinder. Addition to the geometry condition, the resonance frequency depends on the propagation mode. The coupling between the microwave energy with the engine cylinder can be done by different ways [28]:

- Small openings between the resonant cavity and microwave energy source
- Coaxial cable between the source and the cavity with a shape antenna
- Coaxial cable between the source and the cavity with a loop antenna

All modifications to turn a SI engine into a MI engine should need no further change than replacing

the spark plug. For this application a coaxial with a shape antenna is suitable. The coaxial cable works as the transmission line which connects the microwave energy with the engine cylinder through the opening for the spark plug. To calculate the output impedance of the coaxial cable the following equation can be used [28].

$$Z = \frac{60}{\sqrt{\epsilon_r}} \ln \left( \frac{b}{a} \right) \quad (10)$$

Where  $\epsilon_r$  is relative permittivity, and  $b/a$  is the ratio between the total and the centre radius of the coaxial cable. Ideally the input impedance of the resonant cavity is the same like the coaxial cable output impedance, under this condition the maximal power can be transmitted.

### III. CONCLUSION

The potential to reduce fuel consumption and emission is significant by using microwave ignition system instead of one point spark plug ignition. The PhD thesis "Simulation Based A-Posteriori Search for an ICE Microwave Ignition System" from Fung Sun [29] of Glasgow University investigated the optimal antenna shape and length for an HCMI system. Based on this previous work, another project will further improve the performance of the microwave ignition system. To design a realistic computer model of a combustion cylinder, with microwave ignition system, CAD software will be used.

Besides, this project can involve the Glasgow University project "Microwave Ignition System Prototyping for F1 "Green Engines". This project will work together with engine developers to design a prototype with HCMI technology and testing the newly designed system.

### REFERENCES

- [1] Osman Akin Kutlar, Hikmet Arslan, and Alper Tolga Calik, "Methods to improve efficiency of four stroke, spark ignition engines at part load," *Energy Conversion and Management*, pp. 3202-3220, 2005.
- [2] Tae-Hyeong Kwon, "The determinants of the changes in car fuel efficiency in Great Britain (1978-2000)," *Energy Policy*, vol. 34, no. 15, pp. 2405-2415, 2006.
- [3] James A Davis, 4805570, 1989.
- [4] Robert J Schaus, 6608430, 2003.
- [5] P M Najt and D E Foster, "Compression-ignited homogeneous charge combustion," 830264, 1983.
- [6] Jürgen Warnatz, Ulrich Maas, and Robert W Dibble, *Physical and Chemical Fundamentals, Modeling and Simulation, Experiments, Pollutant Formation*, 4, Ed. Berlin : Springer, 2006.

- [7] Kathi Epping, Salvador Aceves, Richard Bechtold, and John Dec, "The Potential of HCCI Combustion for High Efficiency and Low Emissions," *Society of Automotive Engineers*, 2002.
- [8] Rolf Johansson, Daniel Blom, Maria Karlsson, Kent Ekholm, and Per Tunestal, "HCCI Engine Modeling and Control using Conservation Principles," *Society of Automotive Engineers*, 2008.
- [9] Ward and A V Michael, "Combustion in an internal combustion engine," 3934566, 1974.
- [10] Ward and A V Michael, "Reverse stratified, ignition controlled, emissions best time in learn," 3934566., 1991.
- [11] R Schleupen, "Ignition device for high frequency ignition," 6357426, 2000.
- [12] E Schmidt and H O Ruoss, "Device for igniting an air-fuel mixture in an internal," 6918366., 2003.
- [13] E Schmidt, M Thiel, J Hasch, H O Ruoss, and K Linkenheil, "Induction driven ignition," 7204220, 2003.
- [14] K Katsuhiko, A Endo, and J Takezaki, "Ignition system for internal combustion engine," 4446826., 1981.
- [15] Manning and M P, "Plasma ignition system reduces NOx emissions," *Pipeline & Gas Journal*, vol. 222, no. 10, pp. 26-30, 1995.
- [16] F A Pertl and J E Smith, "Electromagnetic design of a novel microwave internal," *Automobile Engineering*, vol. 223, pp. 1405-1417, June 2009.
- [17] Dennis Michael DeFreitas and Albert Migliori, "Ignition methods and apparatus using microwave energy," 5673554, Oct. 7, 1997.
- [18] Dennis Michael DeFreitas, Timothy W Darling, Albert Migliori, and Daniel E Ress, "Ignition methods and apparatus using microwave energy," 5689949, Nov. 25, 1997.
- [19] James E Smith, R Stile, and G Thompson, "Investigation of a radio frequency plasma," in *International Congress & Exposition*, Detroit, 1997.
- [20] Richard Schleupen, "Ignition Device for a High-Frequency Ignition," 6357426, Mar. 19, 2002.
- [21] Volker Gallatz, WO/2005/059356, 2004.
- [22] Volker Gallatz, Nikita Hirsch, and Irina Tarasova, 7770551, 2010.
- [23] Armin Gallatz and Nikita Hirsch. (2010, May) MWI Micro Wave Ignition AG. [Online]. <http://www.mwi-gmbh.com/>
- [24] Minoru Makita and Yuji Ikeda, "Ignition or Plasma Generation Apparatus," 2010/0196208, Jan. 12, 2010.
- [25] N Tran. (2004, Jan.) Microwave ignition for car engines. [Online]. <http://www.microwaveprocessing.com>
- [26] S M Starikovskaia, "Plasma assisted ignition and combustion," *JOURNAL OF PHYSICS D: APPLIED PHYSICS*, pp. 265-299, 2006.
- [27] J Qiao et al., "The development of microwave systems to reduce diesel exhaust emissions," *ICSE*, pp. 570-574, 2003.
- [28] Samuel Seely and Alexander D Poularikas, *Electromagnetics : classical and modern theory and applications*. New York, United States of America: Dekker, 1979.
- [29] Fang Sun, "Simulation Based A-Posteriori Search for an ICE Microwave Ignition System," University Of Glasgow, Glasgow, PhD Thesis 2010.



# The Optimisation of Bio-diesel Production from Sunflower Oil using RSM and its Effect on Engine Performance and Emissions

Abdullah Abuhabaya, Jafar Ali, John Fieldhouse, Rob Brown and Eko Andrijanto

University of Huddersfield  
School of Computing and Engineering  
Huddersfield, UK  
a.abuhabaya@hud.ac.uk

**Abstract**— Bio-fuel production provides an alternative non-fossil fuel without the need to redesign current engine technology. This study presents an experimental investigation into the effects of using bio-diesel blends on diesel engine performance and its emissions. The bio-diesel fuels were produced from Sunflower oil using the transesterification process with low molecular weight alcohols and sodium hydroxide then tested on a steady state engine test rig using a Euro 4 four cylinder Compression Ignition (CI) engine. This study also shows how by blending bio-diesel with diesel fuel at intervals of B5, B10, B15, and B20 can decrease harmful gas emissions significantly while maintaining similar performance output and efficiency. Production optimization was achieved by changing the variables which included methanol/oil molar ratio, NaOH catalyst concentration, reaction time, reaction temperature, and rate of mixing to maximize bio-diesel yield. The technique used was the response surface methodology (RSM). In addition, a second-order model was developed to predict the bio-diesel yield if the production criteria is known. The model was validated using additional experimental testing. It was determined that the catalyst concentration and molar ratio of methanol to sunflower oil were the most influential variables affecting percentage conversion to fuel and percentage initial absorbance.

**Keywords**- bio-diesel; transesterification; response surface methodology; sunflower oil, engine performance and emission

## I. INTRODUCTION

Energy is very important for humans as it is used to sustain and improve their well-being. It exists in various forms, from many different sources. Historically, with economic development, energy needs grew, utilizing natural resources such as wood, fossil fuels, and nuclear energy in the preceding century. However, rising concerns on energy security, economic development, and climate change in the recent past have focused attention on using alternative sources of energy such as bio-fuels. Bio-fuels are the fuels produced from renewable resources, particularly plant derived materials. There are mainly two types of bio-fuels (first generation bio-fuels): ethanol – produced by fermentation of starch or sugar (e.g., grains, sugarcane, sugar-beet, etc.) and bio-diesel – produced by processing vegetable oils (e.g., sunflower, rapeseed, palm-oil, etc.). Another type of bio-fuel is cellulosic ethanol known as second generation bio-fuel, is produced mainly

from wood, grasses and other lignocellulosic materials from renewable sources. Bio-fuels have become a high priority in the European Union, Brazil, the United States and many other countries, due to concerns about oil dependence and interest in reducing greenhouse gas emissions. The European Union Bio-fuels Directive required that member states realize a 10% share of bio-fuels (on energy basis) in the liquid fuels market by 2020 [1]. For bio-diesel production, most of the European countries use rapeseed and sunflower oil as their main feedstock, soybean oil is the main feedstock in the United States. Palm oil in South-east Asia (Malaysia and Indonesia) and coconut oil in the Philippines are being considered. In addition, some species of plants yielding non-edible oils, e.g. jatropha, karanja and pongamia may play a significant role in providing resources. Bio-diesel is derived from vegetable oils or animal fats through transesterification [2] which uses alcohols in the presence of a catalyst that chemically breaks the molecules of triglycerides into alkyl esters as bio-diesel fuels with glycerol as a by-product. The commonly used alcohols for the transesterification include methanol and ethanol. Methanol adopted most frequently, due to its low cost.

Engine performance testing of bio-diesels and their blends is indispensable for evaluating their relevant properties. Several research groups have investigated the properties of a bio-diesel blend with soybean oil methyl esters in diesel engines and found that particulate matter (PM), CO, and soot mass emissions decreased, while NO<sub>x</sub> increased. Labeckas and Slavinskas [3], examined the performance and exhaust emissions of rapeseed oil methyl esters in direct injection diesel engines, and found that there were lower emissions of CO, CO<sub>2</sub> and HC. Similar results were reported by Kalligeros et al. [4], for methyl esters of sunflower oil and olive oil when they were blended with marine diesel and tested in a stationary diesel engine. Raheman et al. [5], studied the fuel properties of karanja methyl esters blended with diesel from 20% to 80% by volume. It was found that B20 (a blend of 20% bio-diesel and 80% petroleum diesel) and B40 (a blend of 40% bio-diesel and 60% petroleum diesel) could be used as an appropriate alternative fuel to petroleum diesels because they apparently produced less CO, NO<sub>x</sub> emissions, and smoke density. Lin et al. [6], confirmed that emission of polycyclic aromatic hydrocarbons (PAH) decreased when the ratio of palm

bio-diesel increased in a blend with petroleum diesel. In general, bio-diesel demonstrated improved emissions by reducing CO, CO<sub>2</sub>, HC, PM, and PAH emissions though, in some cases, NO<sub>x</sub> increased.

The objective of this study was to optimize the production of bio-diesel from Sunflower oil within a laboratory environment and to evaluate its effectiveness through testing using a laboratory engine test rig. The results showed improved engine performance and reduced exhaust gas emissions with levels acceptable to the standard ASTM D6751 (which was correlated to the content of pigments such as gossypol) [7]. A literature search indicated that little research has been conducted using RSM to analysis the optimal production of bio-diesel using vegetable oils. This study intended to make use of the RMS process to maximize the production of bio-diesel (methyl ester in this experiment) from sunflower oil using the conventional transesterification method. In addition to using the RMS for optimizing the methanolysis of sunflower oil it was a desire to develop a mathematical model which would describe the relationships between the variables and so allow yield to be predicted before the production process was finalised.

## II. MATERIALS AND METHODS

### A. Materials

Methanol and sodium hydroxide were purchased from Fisher Scientific (Loughborough, Leicestershire, UK). Sunflower oil was bought from local shops in Huddersfield, United Kingdom. The diesel oil (B0) was obtained for specialist oil suppliers as commercially available diesel is B5. The bio-diesel from sunflower oil was blended at B5 (5% of bio-diesel to 95% of standard diesel by volume), B10, B15 and B20 and evaluated for engine performance and exhaust gas emissions compared to standard diesel.

### B. Fatty Acid Profile

In accord with the approved method of the American Oil Chemists Society (AOCS), the following equation was used to calculate the percentage FFA content of vegetable oils:

$$\% \text{Free Fatty Acid (as oleic acid)} = \frac{T \times M \times 28.2}{W} \quad (1)$$

Where T is Titration value (ml of NaOH), M is Molarity of NaOH (0.025M), and W is Mass of oil sample (g).

## III. EXPERIMENTAL DESIGN

### A. Transesterification process

The presence of NaOH to produce methyl esters of fatty acids (bio-diesel) and glycerol as shown in "Fig. 1". In this study, the reaction temperature was kept constant, at 35°C. The amount of methanol needed was determined by the methanol/oil molar ratio. An appropriate amount of catalyst dissolved in the methanol was added to the precisely prepared sunflower oil. The percentage of the bio-diesel yield was determined by comparing the weight

of up layer bio-diesel with the weight of sunflower oil added.

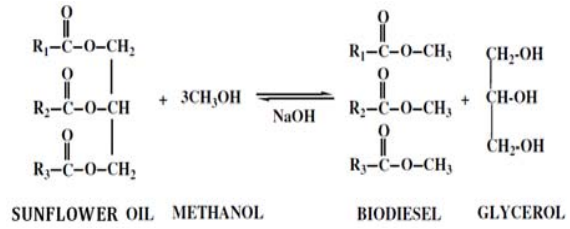


Figure 1. Chemical reaction for sunflower bio-diesel production

Experiments were conducted in a laboratory-scale setup. A 500 ml, three-necked flask equipped with a condenser, a magnetic stirrer and a thermometer was used for the reaction. The flask was kept in the 35°C water bath and stirring speed was maintained at 200 rpm. The reaction production was allowed to settle before removing the glycerol layer from the bottom, and using a separating funnel to obtain the ester layer on the top, separated as bio-diesel.

### B. Optimization process

Optimization of the transesterification process was conducted via a 3-factor experiment to examine effects of methanol/oil molar ratio (*M*), reaction time (*T*), and catalyst concentration (*C*) on yield of methyl ester using a central composite rotatable design (CCRD). The CCRD consisted of 20 experimental runs ( $2^k + 2k + m$ , where *k* is the number of factors and *m* the number of replicated centre points), eight factorial points ( $2^k$ ), six axial points ( $2 \times k$ ), and six replicated centre points ( $m = 6$ ). Here *k* is the number of independent variables, and  $k=3$  should provide sufficient information to allow a full second-order polynomial model. The axial point would have  $\alpha = 1.68$ . Results from previous research [8] were used to establish a centre point of the CCRD for each factor. The centre point is the median of the range of values used: 6/1 for methanol/oil molar ratio, 1% catalyst concentration and 60 min reaction time. "Table (I)" shows the levels used for each factor, and to avoid bias, the 20 experimental runs were performed in random order as shown in "Table (II)". Design-Expert 8.0 software was used for regression and graphical analyses of the data obtained.

TABLE I. INDEPENDENT VARIABLE AND LEVELS USED FOR CCRD IN METHYL ESTER PRODUCTION

Independent Variable	Symbol	Codes and Levels				
		-1.68	-1	0	1	1.68
Reaction Time (min)	(X1)T	43.18	50	60	70	76.8
Methanol/oil Molar Ratio (mol/mol)	(X2)M	4.3	5	6	7	7.68
Catalyst Concentration (wt.%)	(X3)C	0.15	0.5	1	1.5	1.84

The experimental data presented in "Table (II)" was analyzed using response surface regression (RSREG) procedure in the statistic analysis system (SAS) that fits a full second-order polynomial model, "equation (2)". The RSREG procedure uses canonical analysis to estimate stationary values for each factor. Using the fitted model, response surface contour plots were constructed for each

pair of factors being studied while holding the third factor constant at its estimated stationary point. Confirmatory experiments were carried out to validate the model using combinations of independent variables that were not a part of the original experimental design but within the experimental region.

TABLE II. CENTRAL COMPOSITE ROTATABLE DESIGN (CCRD) ARRANGEMENT AND RESPONSES FOR METHYL ESTER PRODUCTION

Run	CCRD component	(X <sub>1</sub> )T (min)	(X <sub>2</sub> )M (mol/mol)	(X <sub>3</sub> )C (wt.%)	Yield (%)
1	Factorial	(-1)50	(-1)5	(-1)0.5	51.09
2	Factorial	(1)70	(-1)5	(-1)0.5	56.60
3	Factorial	(-1)50	(1)7	(-1)0.5	67.94
4	Factorial	(1)70	(1)7	(-1)0.5	72.71
5	Factorial	(-1)50	(-1)5	(1)1.5	54.08
6	Factorial	(1)70	(-1)5	(1)1.5	60.75
7	Factorial	(-1)50	(1)7	(1)1.5	82.93
8	Factorial	(1)70	(1)7	(1)1.5	88.87
9	Axial	(-1.68)43.2	(0)6	(0)1	92.27
10	Axial	(1.68)76.8	(0)6	(0)1	93.17
11	Axial	(0)60	(-1.68)4.32	(0)1	54.63
12	Axial	(0)60	(1.68)7.68	(0)1	94.45
13	Axial	(0)60	(0)6	(-1.68)0.16	26.51
14	Axial	(0)60	(0)6	(1.68)1.8	42.60
15	Center	(0)60	(0)6	(0)1	93.49
16	Center	(0)60	(0)6	(0)1	93.49
17	Center	(0)60	(0)6	(0)1	93.49
18	Center	(0)60	(0)6	(0)1	93.49
19	Center	(0)60	(0)6	(0)1	93.49
20	Center	(0)60	(0)6	(0)1	93.49

$$y = \beta_0 + \sum_{i=1}^3 \beta_i x_i + \sum_{i=1}^3 \beta_{ii} x_i^2 + \sum_{i=1}^3 \sum_{j=1}^2 \beta_{ij} x_i x_j \quad (2)$$

Where  $y$  is % methyl ester yield,  $x_i$  and  $x_j$  are the independent study factors, and  $\beta_0$ ,  $\beta_i$ ,  $\beta_{ii}$ , and  $\beta_{ij}$  are intercept, linear, quadratic, and interaction constant coefficients, respectively. A confidence level of  $\alpha = 5\%$  was used to examine the statistical significance of the fitted polynomial model.

#### C. Engine test setup

The performance of the bio-diesel produced by the transesterification process was evaluated on a Euro 4 diesel engine mounted on a steady state engine test bed. The engine was a four-stroke, direct injection diesel engine, turbocharged diesel, 2009 2.2L Ford Puma Engine as used on the range of Ford Transit vans. The general specification was Bore = 89.9 mm, stroke = 94.6 mm, engine capacity = 2402 cc, compression ratio = 17.5:1, fuel injection release pressure = 135 bar, max power = 130 kW @ 3500 rpm, max torque = 375.0 Nm @ 2000-2250.

Emissions were measured using a Horiba EXSA 1500 system, measuring CO<sub>2</sub>, CO, NO<sub>x</sub> and THC. The test procedure was to run the engine at 25, 50, 75 and 100% engine load over a range of predetermined speeds, 1500, 2200, 2600, 3000 & 3300 rpm. At each of these settings the torque, fuel consumption and emissions were measured for each of the diesels, the standard diesel forming the benchmark.

## IV. RESULTS AND DISCUSSION

### A. Fatty Acid Content Analysis

Since higher amounts of free fatty acid (FFA) (>1% w/w) in the feedstock can directly react with the alkaline catalyst to form soaps, which can then form stable emulsions and prevent separation of the bio-diesel from the glycerol fraction and decrease the yield, it is better to select reactant oils with low FFA content or to reduce FFA in the oil to an acceptable level before the reaction. Nevertheless, the FFA (calculated as oleic acid) content of the sunflower oil used in this experiment was, on average, only 0.13% which was within acceptable levels to be directly used for reaction with the alkaline catalyst to produce bio-diesel [9]. The remaining main factors affecting the transesterification include reaction time, temperature, alcohol/oil molar ratio, rate of mixing, and catalyst concentration.

### B. Response Surface Methodology Analysis

“Table (III)” lists the regression coefficients and the corresponding  $p$ -values for the second-order polynomial model. It can be that the regression coefficients of the linear terms for methanol/oil molar ratio and catalyst concentration ( $M$  and  $C$ , respectively), the quadratic terms in  $M^2$  and  $C^2$ , and the interaction terms in  $TC$  and  $TM$  had significant effects on the yield ( $p$ -value <0.05). Among these,  $M$ ,  $C$ ,  $C^2$  and  $MC$  were significant at the significance level, while  $M^2$  and  $TM$  were significant at the level. Using the coefficients determined from Design-Expert 8.0 software program, the predicted model in terms of uncoded factors for methyl ester yield is:

$$Y_{\text{yield}} = -259.30 - 1.18T + 90.98M + 136.78C - 0.02TM + 0.06TC + 5.99MC + 0.01T^2 - 7.05M^2 - 83.34C^2 \quad (3)$$

TABLE III. REGRESSION COEFFICIENTS OF PREDICTED QUADRATIC POLYNOMIAL MODEL FOR METHYL ESTER PRODUCTION

Terms	Coefficients*	p-value
<b>Intercept</b>		
$\beta_0$	-259.30	0.0001
<b>Linear</b>		
$\beta_1$ (time)	-1.1878	0.6891
$\beta_2$ (molar ratio)	+90.980	0.0001
$\beta_3$ (cat. conc.)	+136.780	0.0003
<b>Quadratic</b>		
$\beta_{11}$ (time)	+0.018	0.6598
$\beta_{22}$ (molar ratio)	-7.052	0.0001
$\beta_{33}$ (cat.conc.)	-83.344	0.0001
<b>Interaction</b>		
$\beta_{12}$ (time and molar ratio)	+0.020	0.0628
$\beta_{13}$ (time and cat.conc.)	+0.06	0.6821
$\beta_{23}$ (molar ratio and cat.conc.)	+5.99	0.0001

\* Because these are calculated values any number of significant figures could be given. However, in the real world an accuracy of 0.01% would be very good so the coefficients are cited to only five significant figures.

The results presented in “Table (III)” suggest that linear effects of changes in molar ratio (M) and catalyst concentration (C) and the quadratic effect  $C^2$  were primary determining factors on the methyl ester yield as these had the largest coefficients. That the quadratic effect,  $M^2$  and the interaction effect MC were secondary determining factors and those other terms of the model showed no significant effect on Yyield. Positive coefficients, as with M and C, enhance the yield. However, all the other terms had negative coefficients. The analysis of variance (ANOVA) revealed that this model was adequate to express the actual relationship between the response and significant variables, with a satisfactory coefficient of determination ( $R^2=0.8142$ ), which indicated 81% of the variability in the response could be explained by the 2nd-order polynomial predictive equation “(3)”. The response surface profile and its contour of the optimal production of yield based on equation “(3)” is shown in “Figs. 2, 3, 4 and 5”, for which the temperature set 35°C, and the rate of mixing was 200 rpm.

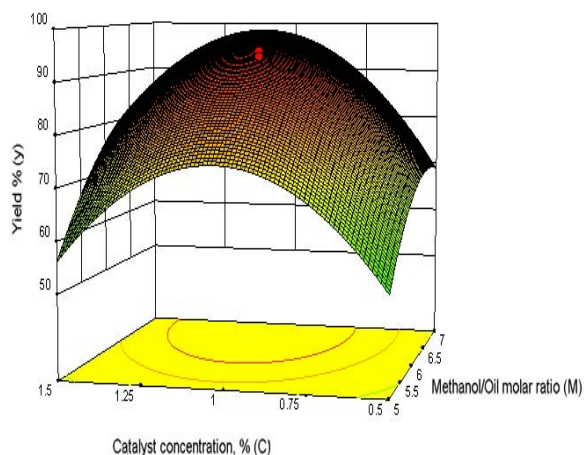


Figure 2. Response surface and contour plot of the effects of methanol/oil molar ratio and catalyst concentration on the yield of bio-diesel with temperature 35°C and reaction time 60 min

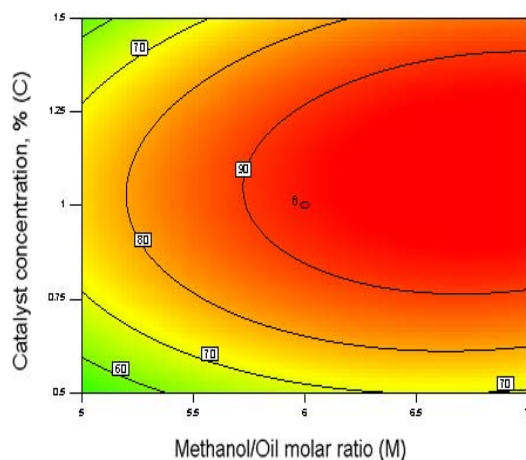


Figure 3. Effect of methanol/oil molar ratio and catalyst concentration on methyl ester production with temperature 35°C and time 60 min

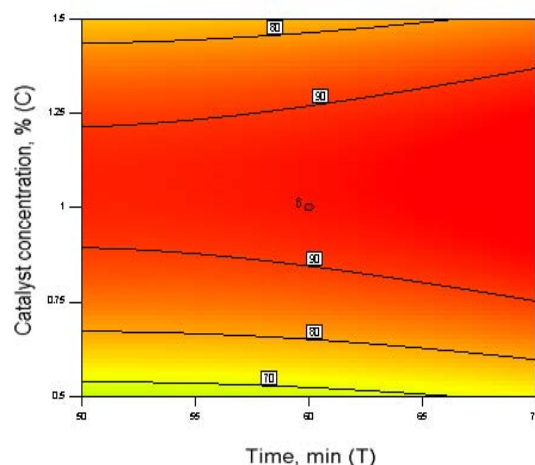


Figure 4. Effect of time and catalyst concentration on methyl ester production with temperature 35°C and methanol/oil molar ratio 7.7:1

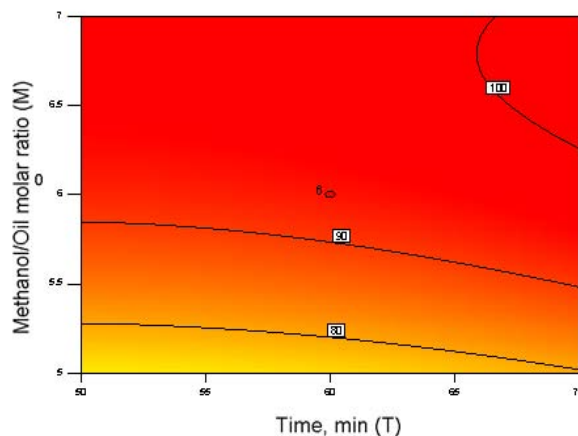


Figure 5. Effect of methanol/oil molar ratio and reaction time on methyl ester production with temperature 35°C and catalyst concentration 1.0%

RSM analysis of the experimental results suggested optimal conditions as: methanol/oil molar ratio, 7.7; temperature, 35°C; time, 60 min; catalyst concentration, 1.0 %; and rate of mixing, 200 rpm. This optimized condition was validated with actual bio-diesel yield of 95%. The decrease of the methanol/oil molar ratio from 7.7/1 to 6.0/1 while keeping the other variable parameters at their respective optimal values produced bio-diesel with a yield of 94%.

### C. Properties of diesel fuel and bio-diesel analysis

The fuel properties of diesel fuel and bio-diesel are presented in “Table (IV)”. The calorific values of the bio-diesel were found using a “bomb calorimeter” to be about 37 MJ/ kg. However, the calorific value of standard diesel fuel was 42.5 MJ/kg, about 13% more than the bio-diesel. The reason for the lower value is because of the presence of chemically bound oxygen in vegetable oils which lowers their calorific values (by about 13 % in this case). It is also shown in “Table (IV)” that the kinematic viscosity of sunflower oil was found to change from 33.72 to 4.53 mm<sup>2</sup>/s at 40 °C, this is a significant change. The initial high viscosity of that oil is due to its large molecular mass in the range of 600-900, which is about 20

times higher than that of diesel fuel, Barnwal et al [10]. The reduction in viscosity during transesterification process reduces the problem associated with using biodiesel in the engine. Density of biodiesel and diesel were determined and found to be about 885 and 845 kg/m<sup>3</sup>, respectively. The flash point of biodiesel was found between 167 and 179°C. Cloud and pour point were also determined and found between -39.7 and 2°C. The properties of the bio-diesel were compared with American Society for Testing and Materials (ASTM) Standard. Most of the fuel properties are found to be in reasonable agreement with ASMT Standard.

TABLE IV. PROPERTIES OF BIO-DIESEL IN COMPARISON WITH THE ASTM STANDARD OF DIESEL AND BIO-DIESEL

#	Experimental results			ASTM D975 Diesel	ASTM D6751 Bio-diesel
Property	Sunflower oil	Bio-diesel	Diesel		
Density(kg/m <sup>3</sup> ) at 15°C	920	885	845	-	-
Kin. Viscosity (mm <sup>2</sup> ) at 40°C	33.72	4.53	2.4	1.9-4.1	1.9-6.0
Calorific value (MJ/kg)	37.26	37	42.54	-	-
Cloud point (°C)	7.2	1	-5	-15 to 5	-3 to 12
Pour point (°C)	-15	-6	-17	-35 to -15	-15 to 16
Flash point (°C)	274	173	76	60-80	100-170
Cetane number (ignition quality)	NA	60	50	40-55	48-60
Iodine number	96.8	NA	NA	-	-

#### D. Engine performance analysis

Sunflower oil itself has relatively low energy content, but the bio-diesel fuel produced from it has a value (about 37.5 MJ/kg, close to that of petroleum diesel; this means that efficiency and output is lower but only by a small percentage. “Figs. 6 and 7” show the curves for power and torque respectively. By simple proportions it the energy content of the blend can be calculated. Energy content of blend = (%diesel x 42.5 + %bio-diesel x 37.5). It can be seen from “Fig. 6” that the loss in power is close to the value predicted. At 20% bio-diesel the calculated power is 41.5 MJ/kg, a decrease of 2.35% compared to petroleum diesel, the measured decrease was about 1.72%.

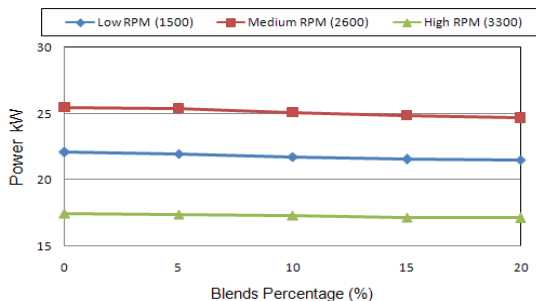


Figure 6. Average power output for different bio-diesel blends

The same trend in the results was seen for torque, there was a progressive decrease in torque as the proportion of bio-diesel in the blend increased, see “Figs. 6 and 7”. The decrease in torque was more apparent than that of the

power, because diesel engines are more focused on torque curves than power curves.

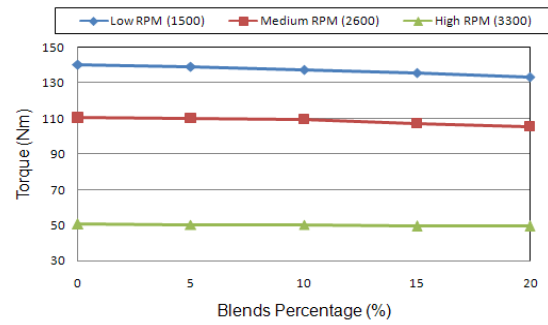


Figure 7. Torque output for different bio-diesel blends

#### E. Engine exhaust gas emissions analysis

As was stated previously the results of bio-diesel blend fuels over the petroleum diesel should show decrease in the emissions of CO, HC, with a slight increase in NOx, and overall similar values for CO<sub>2</sub>. This trend can be seen in “Fig. 8”.

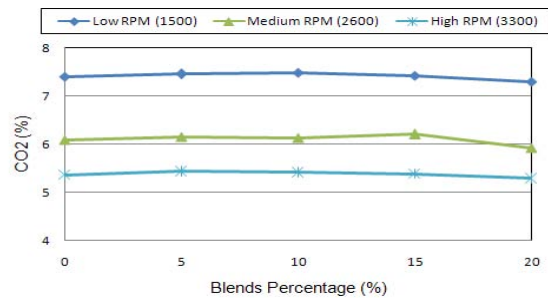


Figure 8. Carbon dioxide emissions for different bio-diesel blends

When bio-diesel is present there is additional carbon, hydrogen and oxygen to be added to the reaction. The resulting problem is seen at B5, this additional carbon caused the emitted CO<sub>2</sub>% to increase. This then falls as the proportion of bio-diesel is increased and a state similar to that for diesel fuel is reached at about B20. Following this trend it is estimated that at higher concentrations of bio-diesel blends (> B20) the CO<sub>2</sub>% emitted would actually be lower than for diesel fuel.

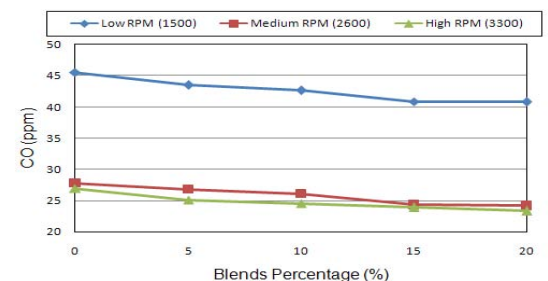


Figure 9. Average CO emission for different bio-diesel blends

The second emission to be analyzed is CO. Carbon Monoxide is present when dissociation is present in the



# *Homogeneous Charge Compression Ignition engine: A Technical Review*

Hammad Iqbal Sherazi\* and Yun Li\*\*

School of Engineering

University of Glasgow

Rankine Building, Glasgow 12 8LT

\*[h.sherazi.1@research.gla.ac.uk](mailto:h.sherazi.1@research.gla.ac.uk) and \*\* [Yun.Li@glasgow.ac.uk](mailto:Yun.Li@glasgow.ac.uk)

**Abstract** — the Homogeneous Charge Compression Ignition (HCCI) engine is the combination of both Spark Ignition (SI-Engine or Otto-Engine) and Compression Ignition (Diesel Engine). It uses homogeneous fuel-air mixtures as the SI-engine does and uses typically high compression ratio that allowed mixture to auto-ignite as the diesel engine does. The lean homogenous mixture in HCCI reduces particulate emissions and lean fuel-air mixture helps auto-ignite which is reducing the emission of oxide of nitrogen NO<sub>x</sub> about 90-98%. It is reported that the HCCI engine offer an increase in fuel efficiency of up to 15-30%, compared with the SI engine. However, HCCI works well only over a relatively narrow operating range, unless engine geometry or operational parameters are adjusted. In HCCI engines combustion is initiated via homogenous charge, and there is no direct in-cylinder mechanism to control auto-ignition. This review paper is focused on HCCI engines performance, challenges, methods to induce combustions, controller design and future direction and trends.

**Keyword** *HCCI engine; Control autoignition; internal combustion engine; Fuel efficient; Emission*

## I. INTRODUCTION

With increasing concerns of finite resources and global warming, researchers in internal combustion (IC) engines are making a tremendous effort to reduce the emission levels and to improve fuel efficiency. While some emissions, such as nitrogen oxides (NO<sub>x</sub>) and soot are immediately harmful to humans and the local environment, carbon dioxide (CO<sub>2</sub>) emissions are receiving more attention due to its increasing effect on climate change. Spark ignition engines operate at well below to their optimum efficiency due to low compression ratios around 8-10 [1] and throttle plate losses used to control air intake. Currently, to improve these losses a new approach is being used called downsizing, in which smaller engines have used to maintain efficiency at high load and consume less fuel [2]. On the other hand Compression ignition engines operate at high efficiency as they use high compression ratios which is around 17-23 [3] but producing soot and NO<sub>x</sub> for which burnout and/or removal can prove problematic. To improve engine efficiency new technologies such as, direct injection gasoline engine, Variable geometry turbocharger (VGTs), cylinder deactivation, variable valve control (VVA), and variable compression (VCR) systems being used. However, after treatment systems, such as particulate filters and selective catalytic reduction (SCR) are typically expensive and do

not solve reduce CO<sub>2</sub> emissions [4]. In order to reduce the CO<sub>2</sub> emissions, engine efficiency must be improved so that per mile CO<sub>2</sub> emissions can be reduced. By altering the combustion mode both issues of improving fuel efficiency and reducing emissions can be addressed simultaneously.

One of the latest strategies is Homogeneous Charge Compression Ignition (HCCI), which seems to be promising solution of all legislative problems. The HCCI engine is the combination of both Spark Ignition Engine and Compression Ignition Engine. It uses homogeneous fuel-air mixtures as the SI-engine does and uses typically high compression ratio that allowed mixture to auto-ignite as the diesel engine does. The combination of these technologies allows HCCI engine to offer several advantages over traditional (IC) engines. In SI engines during combustion, the amount of work (i.e., the load or torque) produced is controlled via throttling process which greatly reduces the engine efficiency. However, HCCI engines generally operate unthrottled and thus do not suffer the efficiency losses observed in SI engines at light loads. As a result, HCCI engine offer an increase in fuel efficiency of up to 30% compared with the SI engine [5].

In addition to fuel economy benefits, HCCI engines also have emission advantages over SI or Diesel engines. In CI engines, the fuel injected into a hot air mass and the fuel starts to auto-ignite shortly after (~1 ms) injection [6]. Therefore, the fuel does not comprehensively mix with the air and there are regions in the cylinder that are extensively fuel rich. These fuel rich regions lead to soot formation [7]. In HCCI engines, the fuel and air are premixed prior to combustion, and thus there are no local fuel rich regions within the cylinder, auto-ignition process occurs throughout the entire combustion chamber with no flame front present. As a result, the combustion temperatures are in the range of 1500K to 2000K temperatures, which is about 500K below conventional diesel and spark-ignition engine technologies [8]. Leads to near zero levels of soot or particulate matter (PM) emission and also produce significantly less nitrogen oxides (NO<sub>x</sub>) than SI or Diesel engines It is reported in [9] that, about 90-98% of NO<sub>x</sub> raw emission reduction can be achieved in HCCI combustion in comparison to conventional combustion.

The most distinctive attribute of HCCI engines, however, is that they have been shown to operate efficiently using natural gas [10], gasoline [11] or diesel [12] fuel, unlike



either SI or diesel engines, both of which require specialized fuels. The fuel tolerance of the HCCI engine is a direct result of the fact that no flame propagation is required, and the only requirement for HCCI combustion is that the fuel ignites at a time close to maximum compression (Top Dead Centre, or TDC) of the piston stroke [13].

## II. CLASSIFICATION OF HCCI ENGINES

Research into HCCI engine began in 1979, when [14] investigated the characteristic of HCCI engines on two-stroke gasoline engines for improved fuel consumption and exhaust emission at part throttle operation. They termed this combustion process as Active Thermo-Atmosphere Combustion (ATAC). In 1983, [15] were able to achieve compression ignition homogeneous charge (CIHC) combustion in a four-stroke gasoline engine. They controlled HCCI with chemical kinetics with negligible influence from physical effects (turbulence and mixing). The literature however becomes dormant until 1989 when Thring [16] revisited the four-stroke engine and explores the effects of exhaust gas recirculation (EGR), intake temperature, and compression ratio; he was the first person to use the acronym HCCI. Throughout the years HCCI has encountered many terminologies in the literature: As, **ATAC** (Active Thermo-Atmosphere Combustion) [17], **LHC** (Lean Homogeneous Combustion) [18], **CIHC** (Compression Ignited Homogeneous Charge Combustion) [19], **AR** (Active Radical Combustion) [20], **HCDC** (Homogeneous Charge Compression Ignition Diesel Combustion) [21], **HIMICS** (Homogeneous Charge Intelligent Multiple Injection Combustion System) [22], **PREDIC** (Premixed Direct-Injection Combustion)[23], **PCIC** (Premixed Compression Ignited Combustion)[24], and **CAI** (Controlled Auto-Ignition) [25].

## III. HCCI ENGINE CHALLENGES

Before HCCI engine implementation into production some challenges need to overcome. This section describes the main challenges, which is been reported so far.

### A. Combustion Phasing Control

One of the principal challenges is controlling the combustion timing in HCCI engines. SI and CI engines have direct in-cylinder mechanism to control combustion timing but HCCI engine lacks in such kind of mechanism. SI engine is controlling combustion event by spark plug and CI engine by fuel injector. In HCCI engine the mixture of fuel and air is premixed and injected in a cylinder at intake stroke cycle same as in SI engine, this is very important that mixture must be premixed before combustion start. In compression stroke when mixture reaches high temperature, which has made cause of combustion same as in CI engine. However, the charges of following combustion are not same as SI and CI engines. This phenomenon actually leads to the main combustion that is dependable on temperature, pressure, concentrations of the participating species and time along

the compression event. The speed of combustion is dependable on chemical kinetics and hereby strongly influenced by concentration of the species, to control this combustion speed and rapid increasing pressure lean mixture possibly used but it is difficult task to control the combustion in HCCI. Combustion timing control is strongly desirable because, if combustion is occurs too early, efficiency suffers and engine damage occurs and if combustion is occurs too late, the chance of misfire increases. Exhaust emission also depends on ignition timing.

### B. HC and CO Emissions

The makeup of HC emission will be different for each hydrocarbon fuel. HC emission also influence by combustion chamber geometry and engine operating condition. When there is not enough oxygen to convert all carbon to CO<sub>2</sub>, some fuel has not burned and some carbon ends up as CO [26]. When the vaporized and compressed air fuel mixture ignited in HCCI engine, and combustion occurs very fast due to fuel vapour close to the wall of the combustion chamber does not burnt completely, this process called wall quenching. This unburned fuel passes out with exhaust gases. This problem increases with engines that misfire [1].

In HCCI engines where peak combustion temperatures in the 1500K to 2000K temperature range, about 500K below conventional diesel and spark-ignition engine technologies [27]. At these low in cylinder temperature of HCCI engine, even the auto-igniting mixture in the centre of the combustion chamber fails to complete the carbon oxide (CO) to carbon dioxide (CO<sub>2</sub>) and the combustion efficiency deteriorates precipitously at lower loads were the lowest temperatures occur.

In HCCI this post combustion CO oxidation becomes inefficient due to the low temperatures.

Fuel and intermediate species cannot react into ultimate products and some hydrocarbons emitted from combustion chamber only partially oxidize to CO [28].

### C. Operation Range

One of the greatest challenges facing HCCI combustion is its limited operating range as compare with SI or Diesel combustion. Operating domain influenced by several factors, including the engine geometry, fuel properties, and auto-ignition initiation process as well. Current research is emphasises on to a light load operation, which is also limited. There are a lot of room for HCCI operation to expand to a higher load but, there is insufficient thermal energy prevents to go beyond the light load operation.

## IV. HCCI IMPLEMENTATION

HCCI execution is totally depended onto an auto-ignition phenomenon. As discussed above, the HCCI engine has no direct in-cylinder mechanism to control combustion. Many indirect mechanisms have been used to control combustion timing in HCCI engine. In [29], the author

arranged a list of all patents which were used to control or made influence in the HCCI for combustion timings, which include: Cooling and heating the intake air, varying mixture temperature, pressure, and equivalence ratio, Using two fuels to vary mixture auto-ignition properties, injecting ozone, pilot igniting with a second fuel, affecting intake temperature with EGR, Controlling intake temperature with residual mass fraction, varying compression ratio (geometric or effective), varying valve timing to affect residual mass fraction or compression ratio, exhaust throttling to control residual mass fraction, controlling combustion chamber wall temperature, controlling mixture temperature via glow plugs, injecting water to control air temperature, and ion current.

In present time, the methods used to induce auto-ignition can be divided in two main categories, chemical control, and thermal control. Chemical control involves the uses of dual fuel, while thermal control involves the change of parameters such as temperature, pressure, and composition at induction phase to control auto-ignition during engine cycle.

Currently, the most popular mechanism among researchers and manufactures are; thermal control through exhaust gas recirculation (EGR), variable valve timing (VVA), variable compression ratio (VCR), and chemical control through dual fuel.

## V. MODELLING

Mathematical engine models are precious tools for predicting and analyzing in-cylinder processes and allow investigation of many engine design alternatives in an inexpensive trend. The engine numerical model can be classified from simple zero-dimensional (Zero-D) or single zone thermodynamic models to complex multi-dimensional (Multi-D) models [34].

However, in most cases (Multi-D) model combine with chemical kinetic model to investigate detail characterise of a HCCI engine, while Zero-D model predominantly used to identify optimum operating points for best performance. Using of the models and engine experiments, researchers can map predictable and actual performance under a variety of operating conditions. In cylinder gas first law of thermodynamics, which implemented into a simple MATLAB or SIMULINK environment, models process.

### A. Chemical Kinetic Model

A chemical kinetic model is very useful for scheming the ignition timing, the rate of burning of fuel and the amount of emissions in HCCI engines when combined with computational fluid dynamic (CFD) codes. There are five categories of chemical kinetic models: detailed, reduced, lumped, skeletal, and global. These have the general characteristics shown in Table 1 [35].

When possibly one of five chemical kinetic models for HCCI engine couple with CFD codes, require tremendous computational resources and large number of configurations need to be consider. To keep computer utilization within practical limits, researchers need to

select Multi-D or Zero-D model with chemical kinetic model in CFD environment. Several modelling studies simulating HCCI conditions using detailed [36], reduced [37], lumped [38], skeletal [39], and global [40] chemical model have been reported for investigating the peak cylinder pressure, predicting ignition time, ignition temperature, combustion duration, indicated efficiency and NO<sub>x</sub> emissions.

TABLE I. CATEGORIES OF CHEMICAL KINETIC MODELS

Category	Description	Species	Reactions
Detailed	The latest "comprehensive" reaction set	100's	1000's
Reduced	A subset of the detailed model	10's	10's-100's
Lumped	A lumped description for larger species	100's	1000's
Skeletal	Employing class chemistry and lumping concepts	10's	10's
Global	Utilizing global reactions to minimize reaction set	<10	<10

### B. Predicting Start of Combustion Model

This model is very important in HCCI analysis that predicts starting of combustion or detecting auto-ignition; in the recent year, researchers have investigated few methods to predict combustion in HCCI and these methods includes knock integral method, temperature threshold method, two-step method, integrated global Arrhenius rate threshold, and shell autoignition method.

**Knock integral method** has been used in SI engine and that basic idea implemented in HCCI engine to detect combustion, but in HCCI, combustion is dependable on concentration of species and in knock integral method species concentration is not included so, this method is not good to use in HCCI. In [41], author has been used knock integral method and reported that, over the range of different value this method is not able to predict accurately. In [42] modified knock integral model has been used, which is much improved as compare to previous model.

**Temperature threshold method** is the simplest among all methods, it consider that combustion has started once the in-cylinder temperature reaches some threshold temperature, which is, calculated on Wiebe function. The authors in [41 and 43] have shown that temperature threshold fail to capture combustion timing over different operating condition. This is also due to the reaction of combustion in HCCI depends on concentration of species not only the in-cylinder temperature.

**Integrated global Arrhenius rate threshold**, it was cleared that the combustion timing is depend on concentration of species as well as temperature of reactant, the researcher consider that combustion initiation point can be modelled with single global reaction rate. Mathematically, this involves integrating

the single Arrhenius reaction rate expression, which is, similar to the each reaction in the model with detail chemistry. Currently, most of the HCCI simulation studies used integrated global Arrhenius rate threshold method that showed great promise in prediction combustion phasing. **Two-step method** also belongs to integrated global Arrhenius rate equation.

**Shell autoignition model** for hydrocarbon fuels, is based on a general eight-step chain-branching reaction scheme with lumped chemical kinetics model using only five representative species in eight generic reactions. This model is aim at prediction of autoignition rather than describing the complete combustion process [44].

## VI. CONTROL STRATEGIES

As illustrated in the previous section, in the real engine operation the parameters, which are used to control the HCCI combustion process, are interacting robustly, finding appropriate parameters that can be used for combustion timing control has been a main control concern for research within the HCCI field. However, to accomplish steady state conditions in the HCCI combustion we have some hurdles in the boundary conditions that have a significant negative impact on the engine performance [30]. While mode transition between HCCI and conventional combustion at different load conditions, especially at lean mixture situation is most challenging task for control engineers. A number of researchers and institutes have established some control strategies, which is, summarize in table number 3.

## VII. CONCLUSION

After reviewing HCCI engines, we obtained an analysis that, HCCI engines has strong potential to improve fuel efficiency than patrol engine, reduce NO<sub>x</sub>, and soot emission than diesel engine. HCCI engine implementation requires three main hurdles; combustion phasing control, HC and CO emission, and operating range, which needs to be overcome for successful HCCI operation. HCCI combustion enormously depended on chemical kinetic.

Computer simulation has become a dominant tool in realizing HCCI and in quest of control strategies for HCCI and has higher flexibility and lower cost compared with real engine experiments. HCCI can be applied to a variety of fuel types and the choice of fuel will have a significant impact on both engine design and control strategies. Single zone model and all parameters that used to study HCCI combustion behaviour, emission, and performance have discussed. Five main control strategies also highlighted, which could be used to maintain, stabilize HCCI operation. Finally, the control of ignition timing, which determines the main combustion phasing and thus has a strong influence on efficiency and operating range of HCCI engine, because early and late combustion can result in heavy knock-like combustion that damages the engine. Therefore, a need of good combustion phasing control is essential to achieve successful HCCI operation.

## VIII. FUTURE DIRECTIONS

The understanding of in-cylinder process in HCCI engine is creditable due to the research, which has carried out during last decade, but some challenges remain. Considering the combustion phasing control; still some research need to be done. HCCI engine complete operating ranges with full load and high speed have to be realized, injection strategies, EGR control, valve timing control and feedback control technologies should developed further. Further, multimode combustion process can be organized and optimized by the control of EGR, VVA, VCR, and dual fuel methods. Therefore, optimization of all control parameters much are needed for improve efficiency and broader operating range. Finally, closed loop control system will construct a bridge between combustion modes and optimize parameter to make HCCI successful.

## ACKNOWLEDGMENT

The first author would like to express his sincere thanks to National University of Science and Technology (NUST), Pakistan, for sponsoring this project.

TABLE II. SUMMARY OF HCCI ENGINE MODELLING AND METHODS USED IN SIMULATION.

Methods	Model	Fuel	Focal point	Model validation	Auto-ignition model	Comments
VVT	SZM[40],[46]	Propane[40] Ethanol and n-heptane[46]	Pressure, combustion timing, IMEP, and Average exhaust temperature via varying IVC and EVO [40]. CA50 with different fuel ratio rate and IVC [46].	Experimental data with fixed operating point[40],[46]	ASS[40],[46]	Well design model with residual fraction value depend on EVO and IVC position [40]. Combustion timing control via different fuel ratio and steady state & transient state data validation been carried out [46].
EGR	TKM[41],[45] SZM[44]	Iso-octane and n-heptane[41] Diesel[44] Propane and gasoline[45]	SOC and CA50 via EGR ratio, temperature and pressure valve at IVC [41]. In-cylinder pressure and temperature [44].	With experimental data at different conditions[41]	MKIM[41] ASS[44] ATS[45]	MKIM model parameters optimize via Nelder-Mead simplex minimization method [41]. Experiments done on diesel

			In-cylinder Pressure via exhaust valve timing [45].	], [44], [45]		engine at different boosting input pressure [44]. Two different fuel been used with negative valve overlap (NVO) strategy [45].
Dual Fuel	SZM[46]	Ethanol and n-heptane[46]	CA50 with different fuel ratio rate and IVC[46].	With experimental data at different operating points[46]	Shell model [46]	combustion timing control via different fuel ratio and steady state & transient state data validation been carried out[46]
VCR	Real Saab engine[47]	Fuel with RON/MON of 92/82[47]	CA50, IMEP, break efficiency and NOx emission via varying inlet temperature and different air to fuel ratio[47].	Experiments done in real time	With pressure sensor and ion current[47]	Observe the effect of inlet temperature and air to fuel ratio on combustion timing. Also cycle to cycle variation considered in both open loop and close loop[47]

TKM = Thermo-kinetic model, SZM = Single zone model, MKIM = Modified knock integral model, ASS = Arrhenius single step, ATS = Arrhenius two step, IVC = Inlet Valve Close, EVC = Exhaust Valve Close, SOC = Start Of Combustion, CA50 = Crank Angle 50 or position where fuel burnt 50%, IMEP = Indicated Mean Effective Pressure, NVO = Negative Valve Overlap

TABLE III. SUMMARY OF CONTROLLER USED TO CONTROL HCCI ENGINE.

Controller	Method	Focal Point	Comments
PID	VGT for varying input temperature [48]. VVA [54]	Controlling CA50 via fuel ratio [48]. Controlling CA50 and IMEP via EVC and injected fuel amount [54].	Gain-scheduled algorithm has adopted and model linearization done with MOESP [48]. Pole-placement algorithm adopted and arbitrarily pole position used, linearization based on set point values [54].
PI	EGR[49] VVA [53]	Controlling CA50 via rebreathing [49]. Controlling IMEP and CA50 via injected fuel amount and IVO [53].	Feedforward controller also used to get optimum equilibrium point on steady state values [49]. MIMO PI controller implemented and feedforward path used to get optimum operating points [53].
MPC	VVA [50]	Controlling CA50 and IMEP via input temperature and IVC [50].	MIMO MPC controller, with Piece-wise linearization has adopted. Cylinder wall temperature model also included in dynamic model [50].
LQR	VVA[52]	Controlling Peak Pressure and End of combustion via molar ratio of inducted gas and IVC [52].	Linearization based on set point values and molar value calculated via EVC/IVO mapping [52].
H2	VVA[51]	Controlling Peak Pressure and End of combustion via molar ratio of inducted gas and IVC [51].	Linearization has done with set points and molar value calculated via EVC/IVO mapping [51].
Neural Network	Chemical Control [55]	Controlling Peak Pressure via injected fuel amount [55].	The ADALINE networks applied with LMS algorithm for network training [55].

VGT = Variable Geometry Turbocharger, MOESP = Multi-variable Output Error State-space model, ADALINE = ADaptive Linear Neuron, LMS = Levenberg-Marquardt Back-propagation Algorithm, IVO = Inlet Valve Open, PID = Proportional Integral and Derivative, PI = Proportional and Integral, MPC = Model Predictive control, LQR = Linear Quadratic Regulator.

## REFERENCE

- Heywood, John B, "Internal combustion engines fundamentals", 1988, Page 58,570.
- Alex M.K.P Taylor, "Science review of internal combustion engines", Journal of Energy Police V 36, 2008, Page 4657-4667
- Widd, A., P. Tunestål, and R. Johansson (2008b): "Physical modelling and control of homogeneous charge compression ignition (HCCI) engines." In Proc. 47th IEEE Conference on Decision and Control. Cancun, Mexico, pp. 5615-5620
- Karl Lukas Knierim, Sungbea Park, Jasim Ahmed and Aleksandar Kojic, "Simulation of Misfire and Strategies for Misfire Recovery of Gasoline HCCI", 2008 American Control Conference, Westin Seattle hotel, Seattle, Washington USA, June 11-13, 2008, Page 3947-3952.
- John P. Angelos, "Fuel effects in HCCI engine", PhD Theses, Massachusetts Institute of Technology, June 2009
- Nicolia, Tm, Carr, D., Weiland, S.K., Duhme, H., von Ehrenstein, O., Wager, C., von Mutius, E. "Urban traffic and pollutant exposure related to respiratory outcomes and atopy in a large sample of children, "Eur. Respir. Journal. 2003, 21, page 956-963
- Novel Combustion Panel," Basic Research Needs for Clean and Efficient Combustion of 21st Century Transportation Fuels", page 15, 2007, This report is available on the web at [http://www.sc.doe.gov/bes/reports/files/CTF\\_rpt.pdf](http://www.sc.doe.gov/bes/reports/files/CTF_rpt.pdf)
- "Journal of power and energy systems volume2 Elkelay, Zhang Yu-Sheng, ALm El-Din Hagar and Jing zhou Yu, Challenging and future of homogeneous charge compression ignition engines: an advanced and novel concept review. 2008" DOI: 0.1299/jpes.2.1108
- N J. Killingsworth, S M. Aceves, D L. Flowers, and M Krstic, "A Simple HCCI engine model for control", IEEE international Conference on Control Applications, Munich, Germany, October 2006
- H Zhao, Z Peng and N Ladommatos, "Understanding of controlled auto ignition combustion in a four-stroke gasoline engine", Proc Inst Mech Engrs Vol 215 Part D, 2001, page 1297-1310
- G.Bression, D.Soleri, S.Savy, S.Dehoux, D Azoulay, H.B. Hamouda, L. Doradoux, N. Guerrassi and N. Lawrence, "A study of methods to lower HC and CO emission in Diesel HCCI", SEA Technical paper series 2008-01-0034, 2008.
- M J. Roelle, N Ravi, J. C Gerdes, "Estimating Thermodynamic State And Ignition in HCCI with Variable Fuel Injection Timing", Proceedings of IMECE 2007, November 11-15 2007, Seattle, Washington, USA.

13. Shigeru Onishi, Souk Hong Jo, Katsuji Shoda, Pan Do Jo, and Satoshi Kato. "Active thermo-atmosphere combustion (ATAC) a new combustion process for internal combustion engines", 1979. SAE 79050
14. Paul M. Najt and David E. Foster. "Compression-ignited homogeneous charge combustion." 1983. SAE 830264
15. R. H. Thring. "Homogeneous-charge compression ignition (HCCI) engines", 1989. SAE-892068
16. S Onishi, S H Jo, K Shoda, P D Jo, and S kato," Active Thermo-Atmosphere Combustion (ATAC) - A New Combustion Process for Internal Combustion Engines", SAE paper number; 790501, 1979
17. Donald J. Pozniak, "A Spark Ignition, Lean-Homogeneous Combustion, Engine Emission Control System for a Small Vehicle", SAE paper number: 760225, 1976
18. P M. Najt and D E. Foster, "Compression-Ignited Homogeneous Charge Combustion", SAE Paper number: 830264, 1983
19. Y Ishibashi and M Asai, "Improving the Exhaust Emissions of Two-Stroke Engines by Applying the Activated Radical Combustion", SAE Paper number: 960742, 1996
20. H Suzuki, N Koike, and M Odaka, "Combustion Control Method of Homogeneous Charge Diesel Engines", SAE paper number: 980509, 1998
21. H Yokota, Y Kudo, H Nakajima, T Kakegawa, and T Suzuki, "A New Concept for Low Emission Diesel Combustion", SAE Paper number: 970891, 1997
22. Y Takeda, N Keiichi, and N Keiichi, "Emission Characteristics of Premixed Lean Diesel Combustion with Extremely Early Staged Fuel Injection", SAE paper number 961163, 1996
23. Y Iwabuchi, T Shoji, K Kawai, and Y Takeda, "Trial of New Concept Diesel Combustion System - Premixed Compression-Ignited Combustion", SAE Paper number: 1999-01-0185, 1999
24. J Li, H Zhao, and N Ladommatos, "Research and Development of Controlled Auto-Ignition (CAI) Combustion in a 4-stroke Multi- Cylinder Gasoline Engine", SAE paper number: 2001-01-3608, 2001
25. "Engineering Fundamentals of the Internal Combustion Engine" by Willard W. Pulkrabek, march 1997. Page number: 278 and 285
26. Novel Combustion Panel," Basic Research Needs for Clean and Efficient Combustion of 21st Century Transportation Fuels", page 15, 2007, This report is available on the web at [http://www.sc.doe.gov/bes/reports/files/CTF\\_rpt.pdf](http://www.sc.doe.gov/bes/reports/files/CTF_rpt.pdf)
27. Dec, J., "A Computational Study of the Effects of Low Fuel Loading and EGR on Heat Release Rates and Combustion Limits in HCCI Engines," SAE Technical Paper 2002-01-1309, 2002, doi:10.4271/2002-01-1309
28. Jeffrey A. Matthews, "Closed-loop control, Variable valve timing control of a Controller-Auto-Ignition engine", PhD thesis, Massachusetts institute of technology, September 2004
29. "Engineering Fundamentals of the Internal Combustion Engine" by Willard W. Pulkrabek, march 1997. Page number: 304
30. Elkelawy, Zhang Yu-Sheng, ALm El-Din Hagar and Jing zhou Yu, "Challenging and future of homogeneous charge compression ignition engines: an advanced and novel concept review", Journal of power and energy systems volume2, 2008
31. Zhao H, Peng Z, Ladommatos N. Understanding of controlled auto-ignition combustion in a four-stroke gasoline engine. Proc Inst Mech Eng 2001;215(Part D):1297-310
32. Xing-Cia Lu, Wei Chen, and Zhen Huang," A fundamental study on the control of HCCI combustion and emissions by fuel design concept combine with controllable EGR. Part 2 . Effect of operating conditions and EGR on HCCI combustion", Journal of Fuel, doi:10.1016/j.fuel.2004.12.015, 2005.
33. Sere Soylu, "Examination of combustion characteristics and phasing strategies of a natural gas HCCI engine", Journal of Energy Conversion and Management, doi: 10.1016/j.enconman.2004.02.013.
34. Zheng, J., Miller, D.L. and Cernansky, N.P. (2004), "A Global Reaction Model for the HCCI Combustion Process," 30th Intl. Symp. on Combust., Poster No.4F1-18.
35. Kelly-Zion, P.L. and Dec, J.E. (2000), "A Computational Study of Effect of Fuel-Type on Ignition Time In HCCI Engines," Proc. Combust.Inst., 28, p.1187-1194.
36. Jincai Zheng, Weiying Yang, David L. Miller and Nicholas P. Cernansky, "Prediction of Pre-ignition Reactivity and Ignition Delay for HCCI Using a Reduced Chemical Kinetic Model".2001-01-1025 @ 2001 SAE
37. Zheng, J., Yang, W., Miller, D.L. and Cernansky, N.P. (2001), "Prediction of Pre-ignition Reactivity and Ignition Delay for HCCI Using a Reduced Chemical Kinetic Model," SAE Paper No. 2001-01- 1025.
38. Zheng, J., Yang, W., Miller, D.L. and Cernansky, N.P. (2002), "A Skeletal Chemical Kinetic Model for the HCCI Combustion Process," SAE Paper No. 2002-01-0423.
39. Jincai Zheng, David L. Miller and Nicholas P. Cernansky," A Global Reaction Model for the HCCI Combustion Process", SAE 2004-01-2950 @ 2004
40. Gregory M. Shaver, Matthew Roelle, J. Christian Gerdes," Dynamic Modeling of Residual-Affected Homogeneous Charge Compression Ignition Engines with Variable Valve Actuation", Journal of Dynamic Systems, Measurement, and Control, SEPTEMBER 2005, Vol. 127 / 374-381
41. Mahdi Shahbakhti and Charles Robert Koch, "Control Oriented Modeling of Combustion Phasing for an HCCI Engine", Proceedings of the 2007 American Control Conference Marriott Marquis Hotel at Times Square New York City, USA, July 11-13, 2007
42. Gregory M. Shaver, J.Christian Gerdes, Parag Jain, P.A. Caton, C.F. Edwards, "Modeling for Control of HCCI Engines", Proceedings of the American Control Conference Denver, Colorado June 4-6. 2003
43. J. Bengtsson, M. Gafvert, and P. Strandh, "Modelling of HCCI engine combustion for control analysis", 43<sup>rd</sup> IEEE conference on Decision and Control, Atlantis, Paradise Island, Bahamas, December 14-17, 2004
44. M. Canova, R.Garcin, S. Midlam-Mohler, Y. Guezennec, and G. Rizzoni, "A control-oriented model of combustion process in a HCCI diesel engine", 2005 American Control Conference, June 8-10, 2005, Portland, OR, USA.
45. N. Jia, J. Wang, K. Nuttall, J. Wei, H. Xu, M.L. Wyszynski, J. Qiao, and M. J. Richardson, "HCCI Engine Modelling for Real-Time Implementation and Control Development", IEEE/ASME Transaction on mechatronics, Vol. 12, NO.6, December 2007.
46. J. Bengtsson, P. Strandh, R. Johansson and B. Johansson, "Hybrid modelling of HCCI engine dynamics-a survey", International Journal of Control, Vol.80, NO.11, November 2007, 1814-1847
47. J. Hyvonen, G.Haraldsson, and B. Johansson, "Balancing cylinder to cylinder variations in a multi-cylinder VCR-HCCI engine", SAE, 2004-01-1897.
48. J. Bengtsson, P. Strandh, R. Johansson and B. Johansson, "Closed-loop combustion control of homogeneous charge compression ignition (HCCI) engine dynamics", Int. J. Adapt. Control Signal Process. 2004; 18:167-179 (DOI: 10.1002/acs.788)
49. C.J. Chiang, and A.G. Stefanopoulou, "Dynamics of Homogeneous Charge Compression Ignition (HCCI) engines with high dilution", Proceedings of the 2007 American Control Conference, Marriott Marquis Hotel at Times Square, New York City, USA, July 11-13, 2007
50. A. Widd, P. Tunestal, and R. Johansson, "Physical modelling and control of homogeneous charge compression ignition (HCCI) engines", Proceedings of the 47<sup>th</sup> IEEE conference on Decision and Control, Cancun, Mexico, December 9-11, 2008
51. G. M. Shaver, J.C. Gerdes, and M. Roelle, "A Two-Input Two-Output Control Model of HCCI Engines", Proceedings of the 2006 American Control Conference, Minneapolis, Minnesota, USA, June 14-16, 2006
52. G. M. Shaver, J.C. Gerdes, and M. Roelle, "Physics-based closed-loop control of phasing, peak pressure and work output in HCCI engines utilizing variable valve actuation", Proceedings of the 2004 American Control Conference, Boston, MA, USA, June 30, 2004 - July 2, 2004
53. M. V. Subbotin, K. L. Knierim, S. Park, A. Kojic, and J. Ahmed, "Modelling and Control of Two Stroke HCCI Engine", ", Proceedings of the 2008 American Control Conference, Westin Seattle Hotel, Seattle, USA, June 11-13, 2008
54. N. Ravi, M.J. Roelle, H.H. Liao, A.F. Jungkunz, C.F. Chang, S. Park, and J.C Gerdes, "Model-Based Control of HCCI Engines using Exhaust Recompression", IEEE Transactions on Control Systems Technology, Vol.18, No.6, November 2010
55. M. Mirhassani, X. Chen, A. Tahmasebi, and M. Ahmadi,"On Control of HCCI Combustion-Neural Network Approach", Proceedings of the 2006 IEEE International Conference on Control Applications Munich, Germany, October 4-6, 2006



# Energy Management System for Tribrid Electric Vehicles

Kary Thanapalan\*, Fan Zhang#

\*Faculty of Advanced Technology, University of Glamorgan, Pontypridd CF37 1DL, UK E-mail: kthanapa@glam.ac.uk

#Faculty of Health, Sport & Science, University of Glamorgan, Pontypridd CF37 1DL, UK

**Abstract** — This paper described the design and implementation of energy recovery systems for tribrid electric vehicles, thereby providing improved performance. Simulation study of control analysis, configuration setup and analysis for better energy management strategies are carried out using simulation tools developed by the University of Glamorgan (UoG), Faculty of Advanced Technology. The simulation tools are used to analyze the effect of operating conditions and energy demand of a hybrid vehicle. Real time implementation of the energy recovery mechanisms are also discussed with references to the University of Glamorgan's tribrid vehicles.

**Keywords-** Powertrain, Energy sources, Energy recovery, Hydrogen vehicles.

## I. INTRODUCTION

The transport sector has the fastest growing demand for energy, yet it is currently almost entirely dependent on the fossil fuels (Zamora *et al*, 2011; Thanapalan and Liu, 2010). UK energy production and consumption for the period of 1970 to 2009 are shown in Figure.1

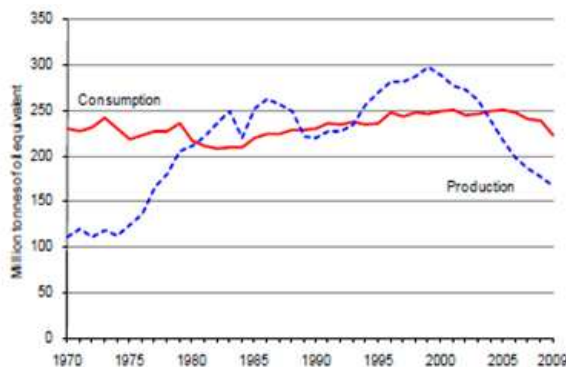


Fig.1. UK energy production and consumption 1970 to 2009 (Mackenzie, 2004)

From Figure.1 it is clear that the energy consumption is growing while production is decreasing. Energy shortage would be further exacerbated by fiscal and economic

measures which may be required to mitigate green house gas (GHG) effects caused by burning fossil fuels. UNFCC-GHG inventory report shows that the UK carbon emissions are significantly high (see Fig.2). In order to reduce the carbon emission, policy makers set some high targets to achieve. The current UK target against 1990 emissions is about 25% and 34% cut in emissions by 2020. Furthermore, 80% cut in emissions by 2050 is also targeted (UNFCC inventory report, 2009).

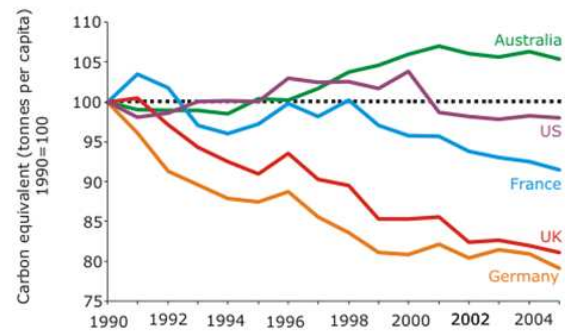


Fig.2. UNFCC green house gas (GHG) inventory

Hydrogen as an energy carrier is of considerable relevance in such circumstances, primarily because it may be produced from several renewable resources, through a number of technological routes (Thanapalan *et al*, 2011a). Hydrogen produced from renewable sources has the flexibility to be used as a clean, safe and convenient transport fuel (Thanapalan *et al*, 2011b). Research and developments have a fundamental role to play in addressing issues of public acceptance, whilst proving the technologies to be employed. To this end, research and development is being carried out to investigate and improve the performance, stability and reliability of the fuel cell based hybrid electric vehicle system, see for example (Andreasen *et al*, 2008; Thanapalan *et al*, 2009a; Hernandez *et al*, 2010; Thanapalan and Liu, 2010). The complexity of the system requires elaborate and innovative studies for proper configuration, component sizing and control system development to fully explore the potential of this advanced technology. A hybrid Fuel Cell (FC) vehicle

contains an energy storage system to provide peak power and capture energy (Rodatz et al, 2005; Viterna 1996; Burke 1996). This usually is a battery or an ultracapacitor pack or a combination of both. Recent studies (Schupbach and Balda, 2003; Thanapalan *et al*, 2009a; Schupbach *et al*, 2003) have shown that the combined battery-ultracapacitor energy storage system can provide better performance and fuel economy.

In this paper, development of energy recovery systems for tribrid electric vehicle alongside energy management system development for performance improvements, are investigated. Essentially and in general terms there are two approaches to energy saving mechanisms; implementation of a better energy management system by providing optimal powertrain topologies for the energy sources and incorporating an energy recovery mechanism. This work addresses both approaches by presenting a powertrain topology for the tribrid electric vehicles (TEV) for better energy management and development of an onboard hydrogen production and storage system for energy recovery. Analyses are carried out using simulation tools developed by the UoG and real time implementations of the energy saving mechanisms are investigated with reference to UoG's tribrid electric vehicles.

## II. SIMULATION TOOLS

Utilizing the hybrid electric vehicle (HEV) system toolbox built in MATLAB/Simulink to facilitate the customized HEV applications, described in (Thanapalan and Liu, 2010), different energy management systems of tribrid vehicle are designed. A quantitative analysis was performed to determine the best powertrain topologies for use in this study, based on simplicity, efficiency, mass and cost etc. Figure. 3 show the chosen topology.

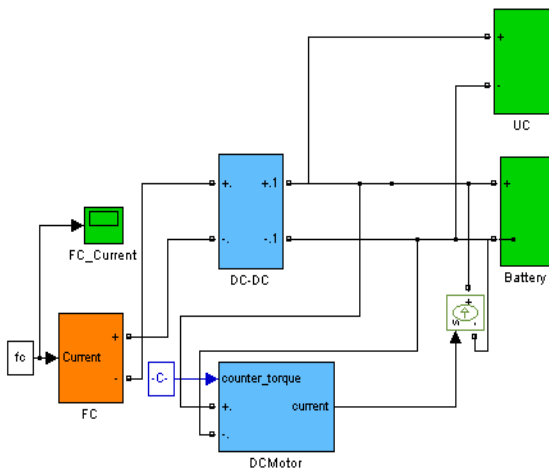


Fig.3.Tribrid energy sources configuration

In addition to the three energy sources; fuel cell (FC), battery and supercapacitor (UC) the configuration consists of a DC motor, and a DC/DC converter. The whole vehicle system consists of hydrogen supplying system, an electric drive and the vehicle body. A typical overall vehicle system configuration is shown in Fig.4. Taking these configurations into consideration, implementation of tribrid electric vehicles is investigated.

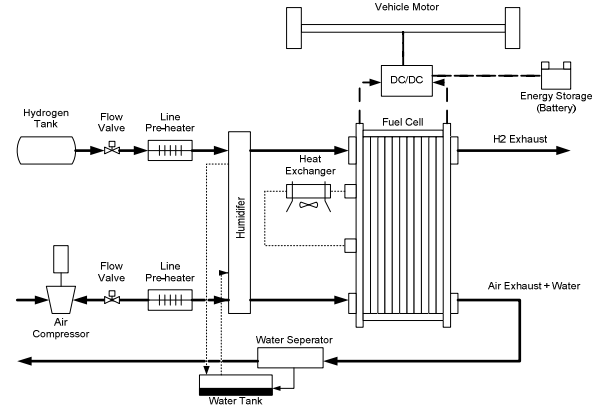


Fig.4. An overall vehicle system configuration

## III. TRIBRID ELECTRIC VEHICLES

The vehicles studied in this paper are the tribrid hydrogen fuel cell vehicles of the University of Glamorgan (UoG). Research efforts at the UoG leads to the production of three tribrid vehicles (see Figures 5, 6 and 7) and associated simulation tools to further investigate and alleviate the problems such as energy management, system configuration, fuel consumption and hydrogen storage etc. these tools are also used to conduct an extensive study for control analysis, and math-based vehicle developments and controller design to improve system performance. In this paper we investigate the energy management and energy recovery mechanism for these vehicles.



Figure.5. University of Glamorgan Hydrogen Scooter vehicle test-bed (Thanapalan *et al*, 2008)

The scooter vehicle is a small scale system to testify the system configuration, reliability and applicability of the powertrain topologies and experimental setup. The test was conducted successfully (Thanapalan *et al*, 2008) and the results leads to the production of the UoG, Faculty of Advanced Technology, Hydrogen Bus (UoGHB). The powertrain of UoGHB consists of a 12kW PEM fuel cell stack developed by Hydrogenics, a 288v, 132 Amp/hr lead-acid battery pack, 375v, 63F Maxwell ultracapacitor, 70kW DC motor, see Figure.8. This is an improved design based on the initial large scale MULE vehicle system, which was developed to testify the same system configuration developed in a scooter vehicle in a large scale. The UoGHB was tested in various conditions to verify its performance while experimental data were collected from the vehicle components for further analysis.



Figure. 6. Large scale MULE vehicle system (Williams and Stevenson, 2010)

It has been observed that, the range of the tribrid fuel cell bus was achieved as 150 miles with top speed 55mph. compared to MULE vehicle this was a significant improvement. In the case the MULE vehicle the range was 37 miles with top speed 40mph. Furthermore, for the UoGHB the maximum power output is 75kW with the maximum power input of 45kW, alongside capacity discharged is 72Ah and energy discharged is 35.6kWh. In contrast for the MULE vehicle maximum power output is 22kW with the maximum power input of 22kW, alongside capacity discharged is 228Ah and energy discharged is 13.9kWh. It is evident that the  $CO_2$  emissions by these hydrogen based vehicles are lower than the existing fossil fuels vehicles. For example, comparison results of the UoGHB,  $CO_2$  emissions with a typical diesel van  $CO_2$  emission for a same distance travelled with a similar environmental and road conditions shows that the  $CO_2$  emissions of the UoGHB is significantly lower than the typical diesel van  $CO_2$  emissions, infact it was observed that the UoGHB,  $CO_2$  output is half of the typical diesel van  $CO_2$  output

(Williams and Stevenson, 2010). It is important to note that these will help to reduce the carbon emission and play a significant role to achieve UK carbon emission reduction targets.



Figure.7. University of Glamorgan Hydrogen Bus (Thanapalan *et al*, 2011a)

The results indicate that the parallel connection of energy sources has improved the system performance significantly. In addition to energy sources configuration for better energy management, the optimal powertrain topology includes a control unit (see Figure.8). There are many different type of control strategies developed to improve the performance of TEV system. It is clear that good performance of the TEV system is closely related to the kind of control used, so a study of different control alternative is justified (Thanapalan *et al* 2009a). A fuzzy logic controller (FLC) is designed for the UoGHB, because the cost of the controller design and implementation is relatively low and it has a high performance/cost ratio (Thanapalan *et al* 2009b). The fuzzy logic controller is used to overcome inherent disadvantages such as uncontrollably large overshoot and large current ripple. FLC is also nonlinear and adaptive in nature and offers robust performance under parameter variations and load disturbances.

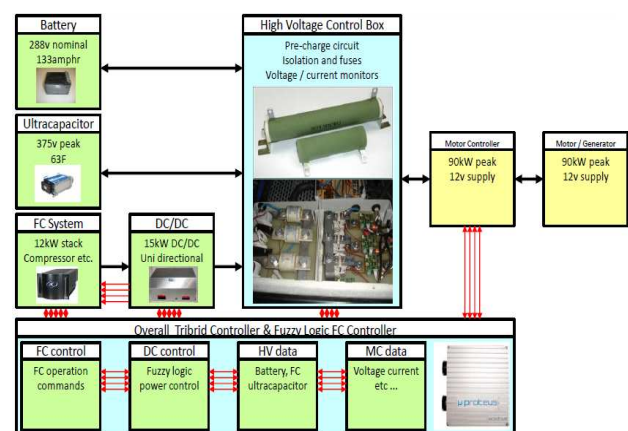


Fig.8. overall powertrain implementation of UoGHB (Williams and Stevenson, 2010)



So, far an optimal powertrain topology for the tribrid electric vehicles (TEV) for better energy management has been established. In the following sections of the paper discusses results from the energy recovery mechanism.

#### IV. ENERGY RECOVERY MECHANISM

##### a. Regenerative braking system

In the past decade, regenerative braking systems have become increasingly popular, recovering energy that would otherwise be lost through braking (Panagiotidis *et al.*, 2000; Cikanek and Bailey, 1995; Cikanek, 1995). In this section, an investigation of the development of a regenerative braking system for tribrid electric vehicle is carried out. The results from the previous research efforts show that, significant amount of energy can be recovered by the integration of the regenerative braking mechanism (Wang and Zhuo, 2008; Dixon and Ortuzar, 2002; Gao and Ehsani, 2001; Schmidt *et al.*, 2000). Further research effort should focus on the development of a systematic and integrated regenerative braking system to recover as much kinetic energy as possible, while maintaining suitable power flows in the other subsystems. To this end, several researchers developed a concept of another energy recovery mechanism (Figure.9) that is still in the research stages is regenerative suspension systems (Electric vehicle news, 2011). This technology has the ability to continuously recover a vehicle's vibrational energy dissipation that occurs due to road irregularities, vehicle acceleration, and braking, and use the energy to reduce fuel consumption.

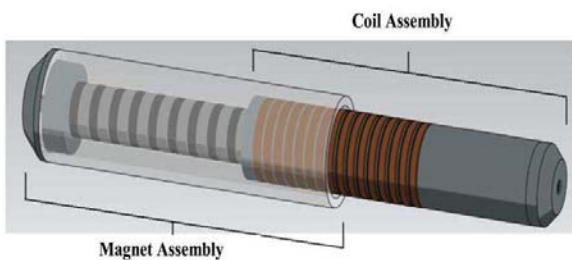


Fig.9. Regenerative shock absorber recovers energy from bumps in the road (Electric vehicle news, 2011)

The utilization of regenerative braking makes the tribrid electric vehicles more attractive by better utilising the on board storage capacity for hydrogen, through raised energy efficiency. Model-based simulation has been used to assist in making an assessment of the benefits from such a system and can be used in exploring further options for improving energy efficiency. In the following section, we will discuss the development of on-board hydrogen production and storage system for hydrogen tribrid vehicles.

##### b. Onboard hydrogen production & storage system

In order to have an efficient economic tribrid vehicle, it is important that the fuel consumption rate should be kept low and that the energy lost should be minimized. Furthermore, overall energy requirements for the vehicle will typically need to be minimized. Many researchers have addresses the issues relating to energy management. However, energy recovery mechanism is an important element in any case of energy usage. In the case of hydrogen based TEV, on-board hydrogen production in addition to energy recovery is highly valued.

The dynamic model of an on-board hydrogen production mechanism from energy otherwise dissipated in suspension damping, contains two interacting subsystems; a generator with an associated electrolyser and a fuel cell stack. The fuel cell and electrolyser models are then incorporated into a system with a suspension energy recovery mechanism (see Fig.10.), used in order to recover energy for subsequent hydrogen production on-board details of the model can be found in (Thanapalan et al, 2011c). In order to maximize the efficiency, the battery is not included due to the fact of its low energy efficiency.

The suspension energy recovery system will recover some of the energy from the active suspension actuator. The recovered energy will drive the electrolyser to produce hydrogen for addition to the stored hydrogen fuel. This subsystem model is implemented in MATLAB/Simulink™ and is parameterized to represent the scooter vehicle simulator (Thanapalan et al 2008; Thanapalan and Liu, 2010), in order to analysis the effect of integrating the on-board hydrogen production mechanism. The results are expected to show that by using this mechanism, energy can be recovered at appreciable rates and hence can contribute additional hydrogen to that stored on-board.

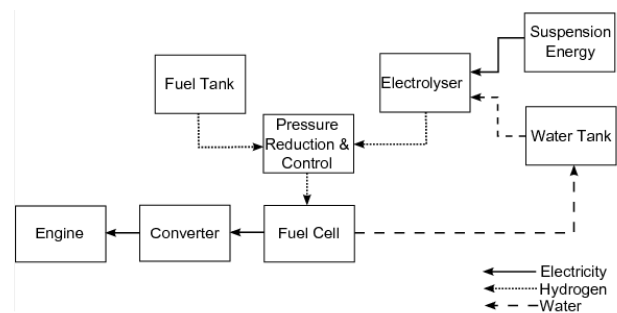


Figure.10. On-board hydrogen production mechanism

The parameters of the suspension system are listed in Table 1. The instantaneous power from the active force is shown in Figure.11 with assumption that the road disturbance is a 2.5mm amplitude sine topology traversed at a speed which will induce a frequency of 10Hz. The negative values in Figure.11 means the active force is against the movement of the suspension and indicates that it allows some power recovery. Ultracapacitor is utilized in order to smooth the fluctuation of the suspension energy recovered.. The energy recovered will be fed to electrolyser for hydrogen production. The peak hydrogen production rate for different disturbance frequencies is shown in Figure 12. And Figure 13 shown the hydrogen production rate for different amplitude of the disturbance.

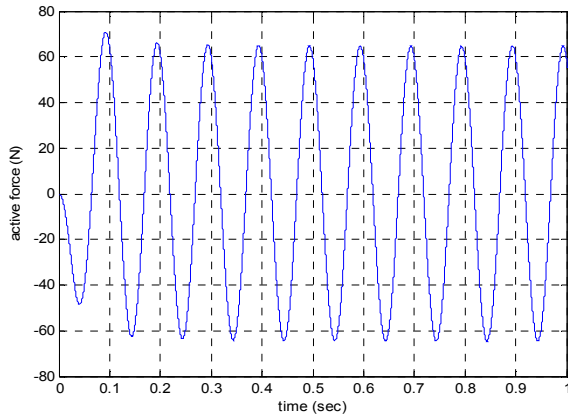


Figure. 11. Active force of the suspension system

TABLE I  
Suspension system parameters

Symbol	QUANTITY	Unit
$m_s$	290	Kg
$m_u$	50	Kg
$k_s$	18000	N / m
$k_t$	210000	N / m

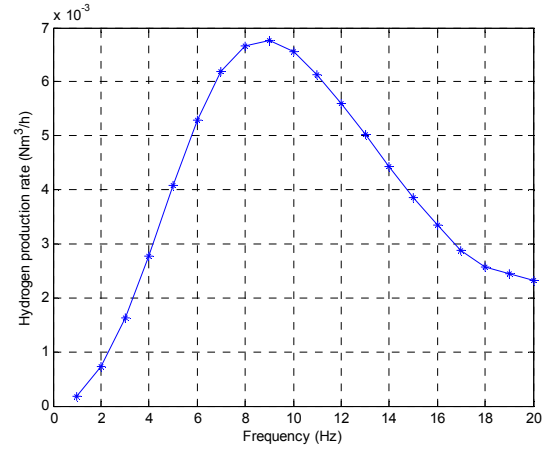


Figure 12. Hydrogen peak production rate with different frequencies

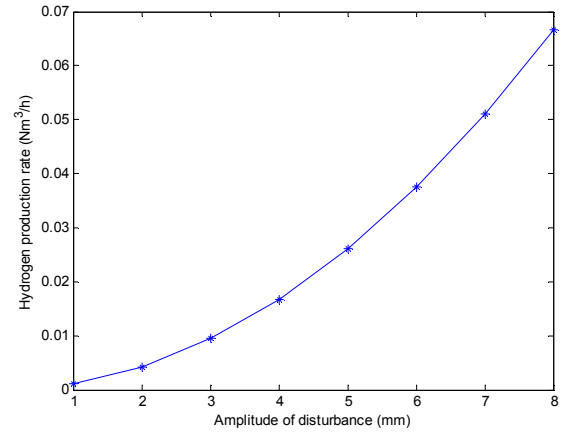


Figure. 13. Hydrogen peak production rate with different amplitude

## V. Concluding Remarks and discussion

The paper describes energy saving mechanisms for tribrid electric vehicles. Simulation study of control analysis, configuration setup and analysis for better energy management strategies are carried out using simulation tools developed by the University of Glamorgan. The HEV simulation toolbox has been updated with the inclusion of a built in demo system of tribrid electric scooter, MULE and bus. These are then used for the development of optimal powertrain configuration for the better energy management. Energy recovery mechanism such as regenerative braking and development of an onboard hydrogen production and storage system are also discussed. It is recommended that an energy recovery mechanism can be used in conjunction with the optimal powertrain topology in order to further reduce the energy consumption of a TEV.



## References

- [1] I.Zamora, J.I.SanMartin, J.Garcia, F.J.Asensio, O. Onederra, J.J. San Martin, V. Aperribay, " PEM Fuel Cells in Applications of Urban Public Transport", In the Proc of Int. Conf. on Renewable Energies and Power Quality (ICREPQ'11), Spain, 2011
- [2] K.K.T.Thanapalan, G.P.Liu, "Modelling and Control of Fuel Cell Hybrid Electric Vehicle Systems", *In the Proc. of UKACC 2010*, Coventry, UK
- [3] W.Mackenzie, "Energy production and consumption" 2004 ([www.woodmac.com](http://www.woodmac.com))
- [4] UNFCC green house gas inventory report, 2009
- [5] K.K.T.Thanapalan, J.G.Williams, G.C.Premier, A.J. Guwy, "Design and Implementation of Renewable Hydrogen Fuel Cell Vehicles", *Renewable Energy & Power Quality Journal (RE&PQJ)*, No.9, May 2011
- [6] S.J. Andreasen, L.Ashworth, I.Natanael, M.Remon, S.K.Kaer, " Directly connected series coupled HTPEM fuel cell stacks to a Li-ion battery DC bus for a fuel cell electrical vehicle", *International Journal of Hydrogen Energy*, Vol.33, pp7137-7145, 2008
- [7] K.K.T.Thanapalan, J.R.Kim, S.J.W.Carr, F.Zhang, G.C.Premier, J.Maddy, A.J.Guwy, "Progress in the Development of Renewable Hydrogen Vehicles, Storage, Infrastructure in the UK : Hydrogen Centre in its early years of Operation", In the Proc of ICICIP 2011, 2<sup>nd</sup> International Conference on Intelligent Control and Information Processing, Harbin, China, 2011, pp738-742
- [8] K.K.T. Thanapalan, G.P.Liu, J.G.Williams, B.Wang, D.Rees, "Review and analysis of fuel cell system modeling and control", *Int. Journal of Computer Aided Engineering and Technology*, vol.1, No.2, pp145-157, 2009
- [9] F.Hernandez, C.Rodriguez, J.L.Hernandez, "Critical analysis on hydrogen as an alternative to fossil fuels and biofuels for vehicles in Europe", *Renewable and Sustainable Energy Reviews*, Vol.14, pp772-780, 2010
- [10] P.Rodatz., G.Paganelli., A.Sciarretta., L.Guzzella., "Optimal power management of an experimental fuel cell/supercapacitor-powered hybrid vehicle", *Control Engineering Practice*, Vol.13, pp.41-53
- [11] L.A. Viterna, "Ultra-Capacitor Energy Storage in a Large Hybrid Electric Bus", In the Proc of 14<sup>th</sup> Electric Vehicle Symposium, 1996
- [12] A.F. Burke, "Electrochemical Capacitors for Electric Vehicles: Technology update and Implementation Considerations", In the Proc of 14<sup>th</sup> Electric Vehicle Symposium, pp.27-36, 1996
- [13] R.Schupbach., and J.Balda., " The role of ultracapacitors in an energy storage unit for vehicle power management", In the Proc of 2003 IEEE Vehicular Technology Conference, Orlando, 2003, pp3236-3240
- [14] R.Schupbach., J.Balda., M.Zolot., B.Kramer., " Design methodology of a combined battery-ultracapacitor energy storage unit for vehicle power management", In the Proc of 2003 IEEE Power Electronics Specialist Conference, Acapulco, 2003, pp88-93
- [15] K.K.T. Thanapalan., G.P.Liu., J.G.Williams., B.Wang., D.Rees., "Tribrid Energy Sources for Electric Scooter Vehicle System", In the Proc of the 14<sup>th</sup> Int Conf on Automation and Computing, London, UK, 2008
- [16] J.G.Williams., P.Stevenson ., "Controlling hydrogen tribrid vehicles", In the Proc of the 6<sup>th</sup> International Conference & Exhibition on Hydrogen & Fuel cells for clean cities: vehicles and buildings, Birmingham, UK, 2010
- [17] K.K.T.Thanapalan., G.P.Liu., J.G.Williams., D.Rees., "Robust Fuzzy Controller Development for A PEM Fuel Cell System", *Int. J. of Advanced Mechatronic Systems*, vol.1, No.3, pp.223-230 , 2009
- [18] M.Panagiotidis., G.Delagrammatikas., D.Assanis., "Development and Use of a Regenerative Braking Model for a Parallel Hybrid Electric Vehicle", *SAE 2000 World Congress*, Detroit, MI, March 2000
- [19] S.R.Cikanek., K.E.Bailey., "Energy Recovery Comparison Between Series and Parallel Braking Systems for Electric Vehicle Using Various Drive Cycles", *ASME International Congress and Exposition*, San Francisco, CA, November 12, 1995.
- [20] S.R.Cikanek., "Electric Vehicle Regeneration Antiskid Braking and Traction Control System", *United States patent*, #5,450.324, September, 12, 1995
- [21] F.Wang., B.Zhuo., "Regenerative braking strategy for hybrid electric vehicles based on regenerative torque optimization control", *J. of Automobile Engineering*, Vol.222, pp.499-513, 2008
- [22] J.W.Dixon., M.E.Ortuzar., "Ultracapacitor + DC-DC Converters in Regenerative Braking System", *IEEE AESS Systems Magazine*, pp16-21, August 2002
- [23] Y.Gao., M.Ehsani., "Electronic Braking System of EV and HEV –Integration of Regenerative Braking, Automatic Braking Force Control and ABS", *Future Transportation Technology Conference*, Costa Mesa, CA, August 2001
- [24] M.Schmidt., R.Isermann., B.Lenzen., G.Hohenberg., "Potential of Regenerative Braking Using an Integrated Starter Alternator", *SAE 2000 World Congress*, Detroit, MI, March 2000
- [25] Electric vehicle news, 'Regenerative shock absorber recovers energy from bumps in the road: <http://electric-vehicles-cars-bikes.blogspot.com> , 2011
- [26] K.K.T. Thanapalan., F.Zhang., J.Maddy., G.C.Premier., "Development of On-board hydrogen production and storage system for hydrogen fuel cell vehicles", *Hybrid and Electric Vehicles Conference*, Warwick, UK, 2011