7TH MIDDLE EAST SIMULATION MULTICONFERENCE

MESM'2005

EDITED BY

Marwan Al-Akaidi

OCTOBER 24-26, 2005

A Publication of

EUROSIS

7TH MIDDLE EAST SIMULATION MULTICONFERENCE

OCTOBER 24-26, 2005

Organised by

EUROSIS

Sponsored by

De Montfort University

IEEE UKRI - SPC

Ghent University

University of Sfax

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

General Conference and Programme Chair

Dr M. Al-Akaidi
De Montfort Univerity,
Leicester, United Kingdom

INTERNATIONAL PROGRAMME COMMITTEE

- Rached Tourki, University of Monastir, Monastir (FSM), Tunisia
- Abdellatif Mtibaa, Universirty of Monastir, (ENIM), Monastir, Tunisia
- Maher Dammak, University of Sfax, (IPEIS), Sfax, Tunisia
- Chokri Rekik, University of Sfax, (ENIS), Sfax, Tunisia
- Mohamed Djemel, University of Sfax, (ENIS), Sfax, Tunisia
- Hafedh Trabelsi, University of Sfax, (ENIS), Sfax, Tunisia
- Raouf Ketata, University of Tunis, (INSAT), Tunisia
- Mohamed M'Saad, GREYC, Université de Caen, ENSICAEN, Caen, France
- Mondher Farza, GREYC, Université de Caen, ENSICAEN, Caen, France
- Olfa Kanoun, University Bundeswehr, Munich, Germany
- Faouzi Derbel, Siemens Building Technologies, Germany
- Ghada Mohammed Amer, Egypt-Benha Higher Institute of Technology, Egypt
- Ahmed Nuseirat, Isra University, Jordan
- Mohammed Zaki Khother, Faculty of Engineering, Jordan University. Jordan
- Musbah Aqil, Applied Science University, Jordan.
- Nabil El-Rosan, Faculty of Engineering, Muata University. Jordan
- Hazim Otum, Faculty of Engineering, JUST, Jordan
- Omer Badran, Al' Balqaa Applied University. Jordan
- Hilal Yousif Al-Rafidain University College, Baghdad, Iraq
- Nabil Derbel University of Sfax, Sfax, Tunisia
- Khaldoon Ghadian, Al-Rafidain University College, Baghdad, Iraq
- Wagah Aljubori, Alasra University, Jordan
- Kasim Al-Aubidy, Philadelphia University, Amman, Jordan
- Jehad Mohamad Alja'am, Quatar University, Qatar
- Adnen Albouchi, University of Monastir, Monastir, Tunisia
- Faysal Mnif, Sultan Qaboos University, Muscat, Oman
- M.Saeed, University College Chester, Chester, United Kingdom

MESM’2005

© 2005 EUROSIS-ETI

Responsibility for the accuracy of all statements in each peer-referenced paper rests solely with the author(s). Statements are not necessarily representative of nor endorsed by the European Simulation Society. Permission is granted to photocopy portions of the publication for personal use and for the use of students providing credit is given to the conference and publication. Permission does not extend to other types of reproduction nor to copying for incorporation into commercial advertising nor for any other profit-making purpose. Other publications are encouraged to include 300- to 500-word abstracts or excerpts from any paper contained in this book, provided credits are given to the author and the conference.

All author contact information provided in this Proceedings falls under the European Privacy Law and may not be used in any form, written or electronic, without the written permission of the author and the publisher.

All the articles published in this Proceedings have been peer reviewed.

EUROSIS-ETI publications are ISI-Thomson and INSPEC referenced

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Ghent University, Faculty of Engineering, Dept. of Industrial Management, Technologiepark 903, Campus Ardoyen, B-9052 Ghent-Zwijnaarde, Belgium or to Philippe Geril, ETI Executive Director, The European Technology Institute, Torhoutsesteenweg 162 Box 4, B-8400 Ostend, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

EUROSIS-ETI Publication

ISBN: 90-77381-21-X

Preface

Dear conference delegate,

I am very pleased to welcome you all to the 7th annual Middle East Simulation Multiconference (MESM05) organized by the European Simulation Society and sponsored by De Montfort University (UK) as well as the IEEE UKRI-SPC.

Our decision to run the 7th MESM in Porto, Portugal jointly with the ESM (European Simulation and Modelling Conference) is based on the fact that we thus make the event more successful by sharing experiences with other delegates from a wider community.

The aim of MESM is to bring people from various parts of the Middle East in contact with colleagues working in Modeling & Simulation from around the world. A number of studies, have found simulation & modeling to be one of the most practical and effective problem-solving techniques presently available, however, there is little practical study literature available to guide those interested in the field. Modelling and Simulation make a powerful combination to improve the systems and organisations of the future.

The Conference highlights recent and significant advances in many research areas of modeling and simulation related to Methodology, Networks Communications, Signal & Image Processing, Biomedical Applications, Industrial Applications, Software Engineering, Neural Network, and Fuzzy Systems.

Your presence at this conference emphasizes an important fact: the challenge and opportunities that surrounding the Modeling and Simulation practice anywhere in the world. We in EUROSIS believe that Modeling & Simulation make a powerful combination to improve the systems and organisation of the 21st century. It will continue to contribute to increase public awareness of the quality of life, and the growing need to improve this through better organizations and systems.

As General Conference Chair, I would like to thank all committee members for reviewing the papers and help in organising this event. Thanks also to Philippe Geril for supporting the conference and giving it his time and input. Last but not least thanks to all authors without the conference would not be a successful conference.

Finally I hope that you all enjoy the event in Porto.

Professor Dr Marwan Al-Akaidi October 4th, 2005
General Conference and Programme Chair
Head of School Engineering & Technology
De Montfort University,
Leicester, LE 1 9BH, UK.
Email: mma@dmu.ac.uk

CONTENTS

Preface ..... **VII**
Scientific Programme **1**
Author Listing **161**

COMPLEX SYSTEMS SIMULATION

Drivers, Benefits and Approaches of Software System Simulation
Christian D. Bodemann and Michael Irvine **5**

A simulation based process for the construction of correct models
Riadh Ben Halima, Khalil Drira and Mohamed Jmaiel **12**

Path following for under actuated marine craft using Line of Sight algorithm
J. Ghommen, F. Mnif and N. Derbel **20**

Statistical Approach for Analysis of Energy System Losses in Jordan
Nabeel Tawalbeh and Ayman Agha **27**

LOGISTICS SIMULATION

Reverse Logistics of end of Life Vehicles Modelling and Simulation of Collection and Disassembly
Gabriella Maria Acaccia, Rinaldo Carlo Michelini and Nicola Qualich **35**

Mixed Automation of a sorting Postal Facility Design and Performance Assembly
Gabriella Maria Acaccia, Rinaldo Carlo Michelini and Nicola Qualich **41**

Stable and robust feedback linearization adaptive fuzzy control of SISO nonlinear systems using Takagi Sugeno fuzzy models
D. Ben Halima and M. Chtourou **49**

NEURAL NETWORKS IN INDUSTRIAL APPLICATIONS

Modeling of Abrasive Fluid Micro Machining by Polynomial Neural Networks and Genetic Algorithms
M. Alitavoli Nariman-zadeh, A. Khalkhali and M. Mehran **59**

Characterization of petroleum wells using fuzzy systems
Rafika Hamza, Khaled Mechichi, Abdellaziz Aboud and Nabil Derbel **64**

CONTENTS

Optimal Control of a Container Crane by A Fuzzy Logic Controller, Robustness and Stability
Najla Krichen Masmoudi, Mohamed Djemel and Chokri Rekik.....**70**

SIMULATION OF PID CONTROLLERS

On the Particularities of Hysteresis Band Widths in Takahashi DTC Strategy: the So-called Focal Speeds
Fatma Ben Salem and Ahmed Masmoudi **79**

PID Control Design of A Blower System using the Pole Compensation Technique
Houda Ben Jmaa Derbel.....**85**

On the design of fuzzy supervised PI controller based on gain and phase margin specifications
L.Tounsi-Rekik and M. Chtourou **90**

PID Neural Tuning for Integrating Processes
R. Chibani and M. Chtourou **95**

PHONE NETWORK SIMULATION

Service-Aware Routing Protocols (SARP): Exploiting mobility with loose delay constraints in ad hoc networks
M. Alchaita, M. Al-Akaidi and J. Ivins.....**101**

Effects of Handoffs on the Performance of Cellular Systems Using Simulation
Nada Al-Khatib and Jamal S. Rahhal.....**106**

Computational Aspects to Calculate the Interference between the High Voltage Transmitters and Communication Cables
Ghada M. Amer.....**111**

IMAGE AND DATA MANIPULATION SIMULATION

Image segmentation based on games theory approach Application to melanoma pattern extraction
Amin Kallel and Imene Khanfir.....**119**

Implementation of 9/7 and 5/3 filters by the *Lifting Scheme*
Dhaha Dia, Mohamed Atri and Rached Tourki.....**126**

Face Recognition using Linear Analysis Techniques and Neural Networks Based Approaches
Ichraq Ben Aouicha, Tarek Tray and Dorra Sellami Masmoudi **134**

DATA MODELLING AND SIMULATION

The Capability of reducing the PAPR by Turbo Coding
Omar Daoud, Marwan Al-Akaidi and Jon Ivins.....**143**

Data Mining for Generating Predictive Models of Automatic Speech Recognition
Ayad T. Al-Zobaydi, Marwan M. Al-Akaidi and Robert I. John **147**

Simulation and FPGA Implementation of a Simple Computer
Kasim M. Al-Aubidy, Ra'ed F. Al-Bader, Ala'a A. Smadi.....**151**

SCIENTIFIC PROGRAMME

COMPLEX SYSTEMS SIMULATION

DRIVERS, BENEFITS AND APPROACHES OF SOFTWARE SYSTEM SIMULATION

Christian D. Bodemann, Michael Irvine
VEGA IT GmbH, Germany
Robert Bosch Straße 7
D-64293 Darmstadt /Germany
Christian.bodemann@vega.de, michael.irvine@vega.de

Keywords: Aerospace, Training, Education, Man-in-the-loop, Real-time simulation

Abstract

Technology is evolving in all technical fields. The efficiency and the productivity of the products is increasing, generation by generation with less machines performing the same or more work. This means that the impact of a failure of one machine is increasing and the financial risk as well. Each generation of machines is intended to reduce the risk by being more reliable. However all this doe not lower the risk if there are humans in the loop. Whenever humans operate or maintain machines the result is very dependent on their knowledge and experience. As more complex the systems get, as more difficult it is to find qualified personal. Once the personal is found it is expensive and increases the costs. Once the costs have increased a new system has to be developed, which should reduce the costs.
This circular dependency is just one of the problems. The development of such complex systems includes normally a high effort on system validation. Meanwhile it is quite common to test and validate the performance and the design of each subsystem by simulation, but it is seen quite often to be to complex to perform a complete system simulation.
The aim of the paper is to demonstrate with three examples what the divers, benefits and approaches are for system simulation. The examples are spacecraft simulators, helicopter maintenance simulations and one training simulator from the chemical industry.

Drivers

Whenever you have humans in the loop, there could be failures. The personal might not be qualified, believing to be qualified and misuse the system or simply make a mistake. Also here the knowledge and experience is essential. The problem is how to train and qualify personal without using and blocking the systems they should operate or maintain. Another problem is the validation of a complete system. It is often very cost intensive and difficult to perform an end to end test. But without a complete system validation the risk is increasing to lose the complete system because of not considered interfaces or incomplete handling procedures. Therefore, as the main drivers for system simulations can be considered:

- Validation: Complex systems need validation and system simulation is a promising option.
- Training: Expensive systems require highly qualified personal in order not to loose or damage the system.
- Cost: Normally hardware solutions are purchased for validation and training. If System validation and training could be even partially done with a software system simulation, additional hardware could be spared and costs reduced.
- Risk: Every complex system is composed of several subsystems supplied by different subcontractors. Each of these has been validated separately. A full system simulation could show already in an early stage interfacing problems and could be later used to evaluate handling procedures without risking the hardware.

Benefits

As mentioned before one approach can be a software system simulation. These are Simulators, which simulate the complete system behaviours and functionality and allow to train and qualify the personal without using the real system. Once a complete system simulator is existing it can be used for various purposes. However as different as the systems can be as different the

solutions can be. One of the largest and extensive examples is the ATV system simulation facility at Les Mureaux / Paris. This simulator uses hardware in the loop. It consists of many different models, which replace missing hardware, stimulate existing hardware and simulate the environment and physical behaviour of the spacecraft. Several different versions exist of this simulator to validate the functionality of the on board software, the functional behaviour of the hardware and the interfaces to external systems. Another version, without hardware in the loop is used for the spacecraft operations training.
But this is just one example. In principal software system simulation can described by three different approaches:

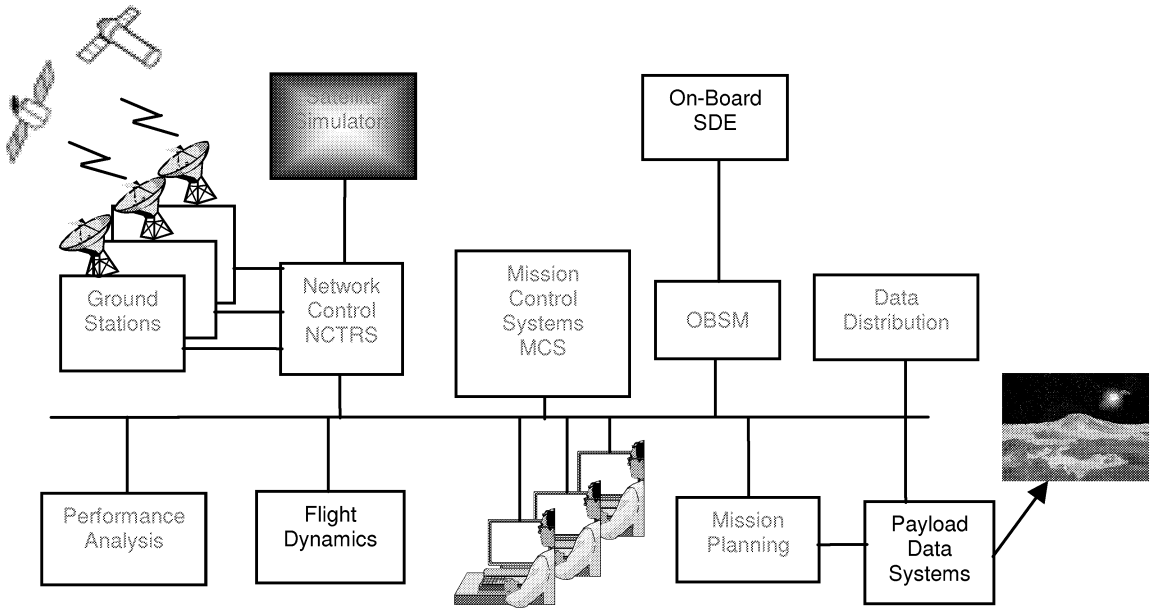
- System Operations Simulation. This kind of simulator normally runs in the background to replace the real system. In this case the interface to the simulator is the real operating system. For example the purpose of a operations spacecraft simulator is to train the control of the spacecraft using the real mission control system.
- Training Simulations. These are mainly used to teach the use of the operation environment. For example how to read and use the instruments in a cockpit of a helicopter. The instruments are fully functional modelled, but it is not the purpose of the simulator to replace a flight simulator, which would be a system operations simulator.
- Computer based Training. These simulators are used for a kind of tutorial training. The simulator is used in the background to simulate behaviour or functionality of a system, while the training system itself teaches knowledge and experience. This is done in a self-training or in a classroom.

In order to explain the three different kinds of system simulations, their drivers and benefits, three examples of existing, successful projects will be described a bit further. It should be noted before that these classifications are not fixed and that there are overlapping between the approaches.

Approaches

System Operations Simulation Example: Venus Express

The Venus Express Spacecraft simulator is a typical example of a System Operations Simulation. As shown in the following picture, the simulator replaces the real spacecraft.



Scope of typical ground segment

The main features of such an approach are:

- It uses the same interfaces to the mission control like the real satellite, accepts and executes the same telecommands as it and provides the same telemetry.
- The functionality of the spacecraft is fully modelled in all details.
- The behaviour of the spacecraft and the simulator is nearly identical.
- In the case of the Venus Express Simulator the on board computer is emulated. This means the original on board software runs in an emulated processor model. This approach allows to upload software patches, download software dumps and debug the on board software. The behaviour of this approach is identical with the real spacecraft data handling system.
- Contingency procedures can be tested and validated as the simulator offers many failure features.

This approach is meanwhile the standard approach for ESA spacecraft operations training. The simulator is used to:

- Train the personal in operating the spacecraft, using the real mission control system six-month prior to launch.
- Validate flight operations procedures prior to training.
- Debug system functionality.

This approach has proven its value to ESA as:

- There is no need for an extra hardware version of the spacecraft. Only the software validation facility is using the electrical validation instance of the spacecraft for short periods.
- It is a highly reliable facility to validate the flight operations procedures, even long after launch.
- It offers a training facility, which is flexible in use and even after years available and usable.
- The whole mission can be flown before the launch of the satellite.
- The ground system can be validated end-to-end using the full chain of subsystems of the ground segment.

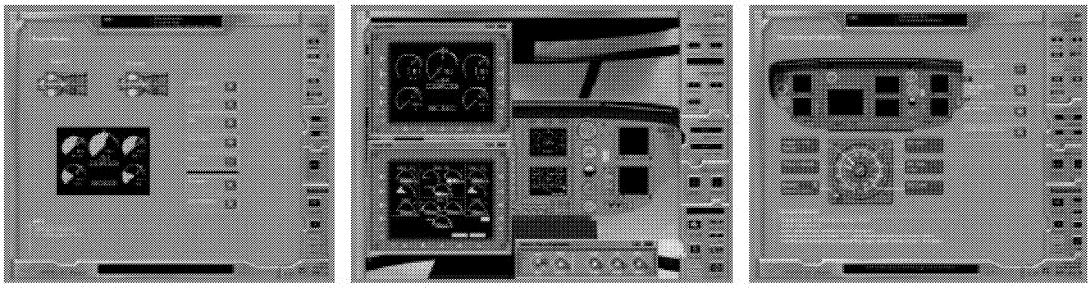
Training Simulations: COUGAR Helicopter Aircrew Training System (CAT)

CAT is a laptop-based training system simulation, which consists of:

- 24 hours of courseware on helicopter familiarisation and operation, including test questions
- An integrated emulation that can be "flown" over a database, using keyboard or emulated cockpit controls.

After using the courseware to build knowledge on these topics, the student can transfer seamlessly to the simulator and use this acquired knowledge to train interactively, performing operational procedures on a variety of the helicopter's subsystems. More than 100 faults can be activated in the emulated helicopter, which the student has to identify and resolve, learning how to operate the helicopter in normal and abnormal conditions. CAT supports initial and conversion training as well as refresher training. The former comprises:

- Sequential use of CBT
- Procedural training guided by CBT audio
- Learning through practical use of the emulation.



The simulator also supports refresher training through the free play mode whilst the CBT provides online help. In addition to introducing components and functionality of the subsystems, the CAT software supports many important training tasks, for example:

- Locating components
- Understanding the subsystems
- Training of subsystem operation
- Developing discipline, proficiency and routine in performing procedures
- Performing all tasks required flying a complete mission.

What is different with CAT?

CAT is a truly integrated training tool that combines learning with training and testing. Exceptional training value results from the immediate transfer of knowledge acquired and seamless transition from building knowledge to applying it on the simulator, with support where needed. The simulator replicates the real helicopter with nearly all subsystems readily available for interactive training. Each student can use it individually, throughout the entire training period (and beyond, if desired) to build and refine their skills. There is no longer a need to wait for the availability of a physical simulator or training aircraft, a laptop is all that the student requires.

Students can operate the simulator under normal and abnormal aircraft states and see if they have acquired adequate proficiency or if there is a training gap. They can even try alternative approaches, and deepen system understanding through a high level of interactivity. As a result, the students become familiar with a high number of fault states and their causes. They acquire well-practised skills, strategy and discipline in their activities, so less training on the full mission simulator or in-flight is necessary.

Advantages

- Integrating CAT in the training concept increases training availability considerably. This is of great benefit to students as it reduces potential time constraints ("training anywhere and at any time")
- Training based on CAT is highly intensive and effective, generating impressive training results
- With CAT designed for self-paced training, larger classes can be managed without compromising training quality
- CAT significantly reduces the direct and indirect cost of training as no expensive equipment, infrastructure or maintenance is necessary to keep the aircrews well trained
- Fully developed and tested, CAT is readily available and can be procured risk-free with customisations that can be easily integrated.

Learning Management System (LMS)

As an option, a powerful LMS is available for integration with CAT to support functions such as:

- Access and account management
- Creation and modification of courses
- Software administration
- Assignment of training tasks
- Compilation of exams
- Automated test evaluation and report generation
- Statistics and overviews.

Additional functionality is available on request. The LMS has an AICC-compliant interface, so the integration of 3rd party media is made easy.

Options and Growth

- Flexibility is provided to tailor the software to meet the individual requirements of various customers, for example:
- Both parts of CAT have built-in flexibility for adaptation to other helicopter configurations, target groups, learning styles, languages and more
- Due to the open architecture, additional procedures or abnormal conditions are easily integrated
- CAT can be configured to use additional training types, e.g. web-based training (courseware part only)
- CAT´s modular design facilitates extension into a part task trainer/training rig or maintenance simulator
- Additional modes of use, namely guided practice and automated procedure tests are provided.

Computer Based Training.: iTiS - Integrated Training, Information and Simulation MERCK

Merck KGaA is one of Europe's leading pharmaceutical companies. It is based in

Darmstadt, not far from the European Space Operations Centre and from VEGA's offices. In late 2000, VEGA approached Merck to find out if the use of software simulations in their training programme was of interest.



Although the Training Department expressed considerable interest, the teams, which operate the plant had a number of reservations. These included the need for time from the plant experts, the cost and the overall suitability of simulations.

In the end, Merck's Training Department decided to ask VEGA to carry out a small pilot project. Support from the European Space Agency's Technology Transfer Programme meant that we were able to increase the scope of what we planned to demonstrate. The project was called iT iS – Merck.

Next to the classroom, which is used for theoretical training, Merck has a laboratory with a number of pieces of equipment to allow hands-on training. This laboratory is used to train not only adults but also apprentices joining Merck from school.

VEGA and Merck selected a multi-purpose training equipment, the Polyapparatur Hastelloy 4000, on which apprentices are trained to carry out a number of procedures also used in larger plants. The Polyapparatur Hastelloy 4000 is used to demonstrate vacuum distillation, heating and cooling techniques, filling and draining procedures, normal pressure distillation and chemical reactions.

For the pilot project, the "filling and draining" procedure was chosen. This is a complex procedure with five main parts and about forty individual steps.

One of the main aims of the iTiS-Merck project was to assess the suitability and potential of simulations in such training programmes. Since feedback from students as well as the instructors was thought to be necessary, it was decided to design and develop a complete training lesson.

The training lesson covered the following:

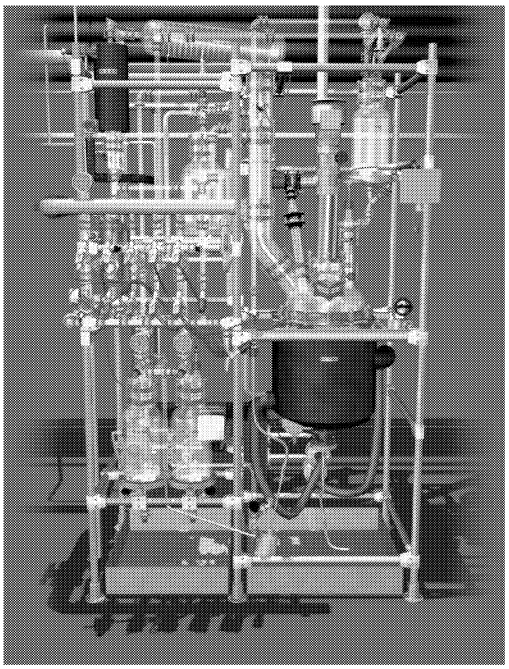
- a general description of the equipment and its purpose;
- a description of the construction of the equipment;
- safety aspects;
- basic processes carried out;
- Learn controls (quizzes, etc.);
- Procedure training.

The lesson was completed in October 2001 and has been used in actual training since that time.

VEGA's motto in technical training has long been "Teach Me, Lead Me, Let Me Practise". Accordingly, the lesson was split into two main parts – an Information Section (to teach the trainees about the equipment) and a Procedure Section (to lead them through the procedure and then allow for practice). Software simulation was used primary in the procedure training part.

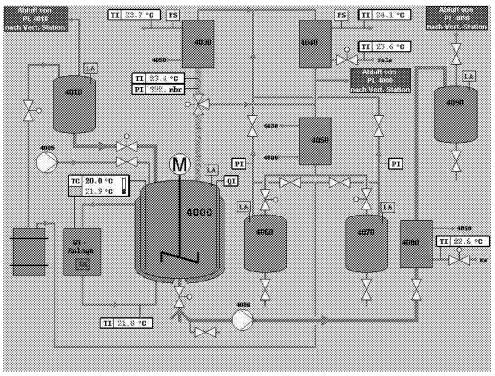
The main features are described below:

- High-quality 3-D graphics were produced at the outset since these were to be used in all parts of the module.
- Animations, photographs, videos, schematics and text were used to describe the equipment and its functions.
- A variety of learn controls were used.
- The SIMSAT simulations kernel was used to run the software models of the equipment in the procedure training.
- The simulation drove all forms of visualisation (i.e. 3-D graphics, 2-D graphics, schematic representations and even photographs).



- An agent was used to explain the procedure to the student. The agent uses all the media available and employs a speech engine to provide audio support.

- The simulation models and the agent behaviour are controlled by a scripting language, enabling the instructors to make simple changes to the training lesson.



The training takes place in a classroom. each student has his own version of the module (running on a two-screen PC configuration). The instructor conducts the lesson using a “Smart Board” connected to his PC.

Conclusion

As the main drivers to use software system simulation can be considered, validation of system, training cost and risks reduction. Software system simulations can be classified into System Operations Simulation, Training Simulations and Computer based Training. In the case of the System Operations Simulation example Venus Express, the example shows clearly the benefit in validation of flight operation procedures and training the operators in advance. Concerning cost and risk reduction it should be noted that the complete spacecraft can get lost in the case of a misuse or a failure in procedures. This would mean a loss of 250 million Euro, spacecraft & launcher loss of scientific results not considered. The reduction of risk using a system simulation facility becomes evident. The reduction of costs using a software solution instead of a hardware system becomes evident when comparing the costs. An extra hardware version of Venus Express would cost about 30 to 50 Million Euro while the software system simulation version is less than 1 Million.

In the case of the Cougar aircrew training simulation the benefits are also evident. The future crew can train the handling and use of the instruments before actually entering the real helicopter. This simulation can and does not replace a hardware flight simulator, but it reduces the time to spend in the hardware simulator significantly. The reduction of risk is given by the fact that this simulation can be run on any PC, which enables every pilot to repeat or re-memorise the functionality and use of the instruments before flight. This can be easily necessary for the case that the pilot was for a longer time not flying this type of helicopter. Considering the reduced flight hours in the military environment due to cost reductions this is quite often the case.

In the case of iTIS the value of a computer based training using a software system simulation in the background is also quite evident. At the beginning of qualification of personal the basics about the functionality have to be trained. In this case it is most beneficial to have a software system simulation in the background, as this is the best way to represent the full functionality of the system without using it. As beginners can easily damage complex systems by mistake, they can get in such a training environment their first experience and can understand the system without the risk of a damage or impact on the production. Not only the risk is reduced by this, also the cost. The costs for such a case can be easily estimated when assumed that the production chain is blocked

for just one hour. In the case of a complex system the costs of one hour are normally far higher than the cost for a CBT. Software System Simulation can be therefore considered as quite beneficial for complex systems.

In the case of the Spacecraft System Simulation this approach demonstrated it's value already several times and for the future it is planed to develop it even further and develop a spacecraft operations training centre. This training centre will allow to train engineers in all functions of spacecraft operations without using a real control centre and a real satellite.

References

[1] DASIA Conference 2005:

Venus Express: Technology is not everything.
Christian D.Bodemann, Joachim Ochs, Ed Curran, Carol Quirke

[2] DASIA Conference 2005:

Rosetta Simulator: Long term missions and simulator maintenance challenge.
Christian D. Bodemann, Joachim Ochs

[3] VEGA IT GmbH Internal Publication 2004

COUGAR Helicopter Aircrew Training System (CAT)

Dr. Matthias Purpus

[4] VEGA IT GmbH Internal Publication 2003

iTiS - Integrated Training, Information and Simulation MERCK

Dr. Peter Fritzen

A simulation based process for the construction of correct models

Riadh Ben Halima, Mohamed Jmaiel
University of Sfax
National School of Engineers
B.P.W, 3038 Sfax, Tunisi
{Riadh.BenHalima@isetma, Mohamed.Jmaiel@enis}.rnu.tn

Khalil Drira
LAAS-CNRS
avenue de Colonel Roche
31007 Toulouse Cedex 4, France
Khalil@laas.fr

Abstract

This paper provides a graphical simulator that enables to conceive correct architectures based on their formal specification. The simulator initially accepts (as an input) a Z specification already validated using Z/EVES tool. Then, it generates graphical entities -according to the UML notation- that represent software's components and their connectors. Thereafter, the user may generate several architectures instances by adding components and connections between them. Architecture instances can be updated by destroying components/connections or by modifying their interconnections. The user actions are checked through the formal specification of the architectural style. Architecture constraints can be adapted according to the users needs, until a complete specification is obtained.

1 Introduction

Different terminologies and understandings of requirements engineering make good communication and good requirements specifications a hard task. Incompleteness, ambiguity, inconsistencies and vagueness are common problems encountered when eliciting and specifying requirements. Furthermore, the validation process of a software architecture is undoubtly an indispensable part of requirements engineering. Indeed, validation is the process of checking whether requirements specifications meet the intentions and expectations of the users.
The construction of applications with dynamic architecture results from the addition or the removal of components and connections between them, and from the modification of interconnections as described in the specification. Designing, this kind of applications is considered as a complex task which requires a rigourous specification and a simulation stage. Indeed, this specification should describe, in a precise way, the architectural style (static aspect) as well as the operations of reconfiguration (dynamic aspect). Moreover, the simulation process should verify the behaviour of the specification .The main advantage of using formal techniques lies in formally verifying that the specified operations of reconfiguration preserve the required architectural style during the application evolution. Hence, it is important to be able to define correctly the constraints which govern the changes of software architectures. Our approach aims to support the process of validation based on simulation of system (behaviour) models that are derived from initial formal specifications. However, the benefit of software models is determined by the capabilities of the corresponding simulation tools.

In this paper, we use formal methods for describing dynamic software architectures. We use the Z notation [15] to specify architectural styles and reconfiguration operations. The use of the formal approach allows us to rigourously reason about architectural properties and, particularly, to prove that the architectural style is respected by the reconfigurations, and hence preserved during the application execution. In order to consolidate this formal approach, we design and implement a graphical tool enabling

to apprehend, also by nonspecialists, the formal Z specification of an architecture. This graphical tool allows to simulate visually the behaviour of a component-based application using the UML2.0 notation [6], which respect the specified architectural style. Across its graphic interface, the simulator allows to the designer, who does not necessarily master formal specifications, understanding the behaviour of the system to build. Moreover, it permits to generate many architecture instances. Whether we conceive an erroneous architecture (the initial Z specification allows this configuration), we give the possibility to modify the Z specification (adding/removal constraints, component and connection) and to reload the specification in order to resume simulation. In the end of the simulation stage, we synthesize a correct specification. Furthermore, it helps to check the completeness of the architectural constraints specified in Z.

In order to achieve this task in a good way, we integrate our simulator with the Z/EVES tool [14] which allows edition, syntax and type checking and reasoning (by proving theorems) about software architectures specified in Z. Moreover, it helps to check the consistence of the formal specification of the architectural style.

In this paper, we emphasize the conception and the implementation of this simulation environment as well as its integration with the Z/EVES tool. We illustrate our approach with a correct architecture specification (proved and simulated)of a Patient Monitoring System (PMS).

This paper is organized as follows. Section 2 gives a survey on the related work. Section 3 introduces the adapted approach of the dynamic architectures description using Z and its application to the PMS. Section 4 presents the features and the design of our simulator. Section 5 describes the different steps of our simulation process. Finally, Section 6 provides some concluding remarks.

2 Related works

In the context of requirements engineer, simulation is defined as the visualisation of the behaviour of a system model. The modelling language used to express the system model must rely on defined execution semantics. Based on the semantics, a simulator tool can visualize the model by direct interpretation. However, the benefit of software models is determined by the corresponding simulation tools. As a consequence, several researches are focused in this axis and many tools are developed in order to validate the requirements specification. Among the industrial solutions we quote: ObjectGeode [2] is a a commercial toolset dedicated to analysis, design, verification and validation through simulation and testing of real-time and distributed applications. It supports a coherent integration of complementary object-oriented and real-time approaches based on the UML, SDL (Specification and Design Language) and MSC (Message Sequence Charts) standards languages. ObjecTime Developer [3] is a graphical modeling environment for the object-oriented design (with UML) and simulation of real-time systems. It allow code generation for C++. IBE PACE [1] is a widely used object-oriented simulator-development-system based on attributed stochastic Petri-nets with time and fuzzy modeling. It allows to model, simulate, analyze and optimize commercial, technical and logistics processes in a comfortable tool environment. But it doesn't provide any proving method's. All of those simulators use UML1.x. Therefore, they don't support the simulation of a component-based model. Besides, most of them don't use a formal specification as an input for the simulation stage.

Currently, the Unified Modelling Language (UML) [6] is accepted as a standard for designing software applications. Its a set of notations helps system designers to represent their solutions in a way that is expressive and yet easy to understand. However, UML is generally criticized for the lack of means for verifying and validating the designed models. Contrary to UML, our approach privileges the use of formal methods for the specification, the standard notations for the presentation, and the dialogue with the user. Its advantage appears in its capacity to produce a formal model presented to the user according to the UML visual notation during the simulation process. So, on one hand, contrary to an approach using only the formal, we suppose that the user of our environment does not need to be an expert in formal methods. On the other

hand, we do not reduce the rigourous capacity of our model by limiting us to the informal side, using only the UML notation.

Several research works choose the use of formal methods for specifying dynamic software architectures. Among these works, we mention the CHAM [16] approach which is based on grammars and rewriting rules. This approach uses reactions and evolutions to formalize, respectively, architectural styles and reconfigurations. We note, however, the absence of tools for validating or simulating specifications developed according to this approach. Other approaches, like LEDA [5], Wright [4] and Darwin [11] make use of process algebra to specify dynamic software architectures. Wright presents the structure of an architecture as a graph and specifies the system behaviour and reconfigurations using a variant of CSP. LEDA and Darwin use the π-calculus to describe dynamic architectures. LEDA supports prototype execution using a π-calculus analysis tool : the *Mobility Workbench* [17]. Another widely recognized approach is RAPIDE [10], in which a software architecture consists of a set of module interfaces interconnected according to a set of connection rules and communication constraints. These prominent approaches, although they propose rigorous solutions for designing software architectures, they lack tools for simulation and animation of software architectures. RAPIDE is one of the few formalisms which enable the simulation the program behaviour in terms of causal ordering of events.

On the other hand, there are several works which tried to animate or to simulate Z specifications, namely, Jaza (Just Another Z Animator), PiZA (A Z Animator which is implemented in Prolog) and ZANS (Z ANimator System). The Jaza tool [13] is used to evaluate the schemas of a Z specification. However, it contains much more disadvantages than advantages. Among the many disadvantages, we can appoint : the absence of the treatment of all the Z language constituents (e.g. generic definitions and axiomatic declarations), the impossibility of declaring the functions or relations infixed, prefixed and postfixed. The PiZA tool [7] allows the conversion of restricted parts of a Z specification into Prolog and their execution. It is hard to use because it requires the installation of other software and contains at least two translation stages. The ZANS tool [8] permits the evaluation of expressions and predicates and the execution of operation schemas. It allows animating finite subsets of Z specifications. However, this tool possesses a weak expressive capacity. Indeed, the state and the initialization schema are not dealt with.

3 Architectures specification with Z

Our work is based on the approach [9], which relies on the integration of graph grammars within the Z notation to describe the static and dynamic aspects of software architectures. Its advantage is that it is expressive enough to deal with the static aspects, namely the component types, the connection types and the architectural properties, as well as the dynamic aspects of an architecture in terms of reconfiguration operations. This approach aims, mainly, at guaranteeing the preservation of architectural properties during the system evolution. According to this approach, the architectural style of a software system is described using a Z schema as follows :

Style

Component types
Connection types

Architectural properties

Schema 1: *Architectural style*

To give more details on our approach, we present in this section a simple example. The Patient Monitoring System (*PMS*), which was used to illustrate works of [12] and [18]. It is a software system which let the nurses (of type *NURSE*) controlling their remote patients in a private clinic. A controller (of type *BED_MONITOR*) is attached to the bed of the patient. For each service of the private clinic (pediatric, cardiology, maternity...), we associate an event service (of type *EV_SERV*) to manage the communications between nurses and bed monitors. For each bed monitor, the responsible nurse periodically requests patient data (for example, blood pressure, pulse and temperature) by sending a request to the event service to which it is connected. This service transmits the request to the concerned bed monitors. When a patient's state is considered to be abnormal, its corresponding bed monitor raises

an alarm to the event service to which it is connected. Then, this service transmits the signal to the responsible nurse. The architectural style is represented by schema 2.

$$
\begin{array}{|l}
\hline
PMS \\
CN : \mathbb{F}\, NURSE \\
PB : \mathbb{F}\, BED_MONITOR \\
ES : \mathbb{F}\, EV_SER \\
pull_NE : NURSE \leftrightarrow EV_SER \\
pull_EB : EV_SER \leftrightarrow BED_MONITOR \\
push_BE : BED_MONITOR \leftrightarrow EV_SER \\
push_EN : EV_SER \leftrightarrow NURSE \\
\hline
\#ES <= 3 \\
\forall s : ES \bullet \#\{(\mathrm{ran}(\{s\} \lhd push_EN))\} \leq 5 \\
\forall b : ES \bullet \#\{(\mathrm{ran}(\{b\} \lhd pull_EB))\} \leq 15 \\
\forall x : PB \bullet \exists_1 y : ES \bullet (x,y) \in push_BE \\
\forall x : PB \bullet \forall y : ES \bullet \exists z : CN \bullet (x,y) \in \\
push_BE \Rightarrow (y,z) \in push_EN \\
\forall x : CN \bullet \exists_1 y : ES \bullet (x,y) \in pull_NE \\
\hline
\end{array}
$$

Schema 2: *PMS schema*

In the first part of schema 2, we develop a declaration part which specifies component types as well as connection types being able to exist between them. In this specification, *CN*, *PB* and *ES* represents respectively the nurses, the bed monitors and the event services. This part integrates also relations representing the communication links between the components.
In addition to the architectural style constraints, an application can have specific properties which must be satisfied during the evolution of its architecture. We will take some properties of the *PMS* system, which are formulated in the second part of schema 2, such:

- The system must contain a maximum of 3 services.
- A service contains a maximum of 5 nurses and 15 patients.
- A patient must be always affected with only one service. This later must contain at least one nurse to take care of this patient.
- A nurse must be connected to only one service.

According to the approach presented in[9], the dynamic part of an architecture is described using Δ (Delta) Z schemas (cf Schema 3). Indeed, each Δ schema represents a reconfiguration operation.

$$
\begin{array}{|l}
\hline
Operation_Name \\
\Delta Style \\
Par_1?;Par_2?;....;Par_n? \\
\hline
Pre\text{-}conditions \\
Post\text{-}conditions \\
\hline
\end{array}
$$

Schema 3: *A Delta schema*

Where $Par_i?$ denotes an input of the operation. The PMS specification presents different operations of reconfiguration that make possible system evolution. The application of a reconfiguration is then performed as follows: if the functional and structural constraints are satisfied, then the actions can be executed. In the following, we present some of them :

- Insertion of an event-service: this rule inserts an instance of a component of type *EV_SER* provided that the system does not already contain three event services according to the first predicate in the PMS schema. This rule is expressed by the following schema *insert_ES*:

$$
\begin{array}{|l}
\hline
insert_ES \\
\Delta PMS \\
x? : EV_SER \\
\hline
\#ES < 3 \\
ES' = ES \cup \{x?\} \\
\hline
\end{array}
$$

Schema 4: *Insertion of an event-service*

- Insertion of a nurse: this rule inserts an instance of a component of type *NURSE* and connects it to an event service (that exists in the system) with the relation specified in the declaration part of PMS schema. We express this rule with the following schema *insert_CN*:

$$
\begin{array}{|l}
\hline
insert_CN \\
\Delta PMS \\
y? : NURSE \\
\hline
CN' = CN \cup \{y?\} \\
\exists x : ES \bullet push_EN' = push_EN \cup \{(x \longmapsto y?)\} \\
\wedge\ pull_NE' = pull_NE \cup \{(y? \longmapsto x)\} \\
\hline
\end{array}
$$

Schema 5: *Insertion of a nurse*

- Removal of a patient: when a patient leaves the system, the corresponding bed monitor is disconnected from the associated service and it is removed from the system. This rule is expressed with the following schema *supp_PB*.

supp_PB

ΔPMS

$z! : BED_MONITOR$

$\forall x : ES \bullet push_BE' = push_BE \setminus \{(z! \longmapsto x)\}$

$\wedge\ pull_EB' = pull_EB \setminus \{(x \longmapsto z!)\}$

$PB' = PB \setminus \{z!\}$

Schema 6: *Removal of a patient*

The adopted approach aims to facilitate the task of the developer by offering to him an operational technique of specification. This technique, which is easy to apprehend, provides a rigorous reasoning to analyse and to verify the evolution of a software architecture. We used the Z/EVES tool [14] to edit and check the syntax and the types of the PMS specification as well as to prove that the specified reconfiguration operation preserve the architectural style. Indeed, if we part from a configuration which satisfies the architectural style and we apply an operation of reconfiguration to the latter, we obtain a new configuration which also satisfies the architectural style. Thus, it helps to check the consistence of the specification of the architectural style.
The following phase debates the visual simulation of a Z specification. It will be detailed in the next section.

4 The simulator : presentation and features

The simulator is a generic application which simulates the functioning mode of component-based applications. It represents components and connections using UML 2.0 conventions. Moreover, it provides the possibility of adding and/or of removing components and connections. Each reconfiguration action (adding/removing) proceeds while respecting its description within the Z notation, according to the approach presented in Section 3. The main interface of the simulator (figure 1) contains menus, a toolbox and a drawing panel.

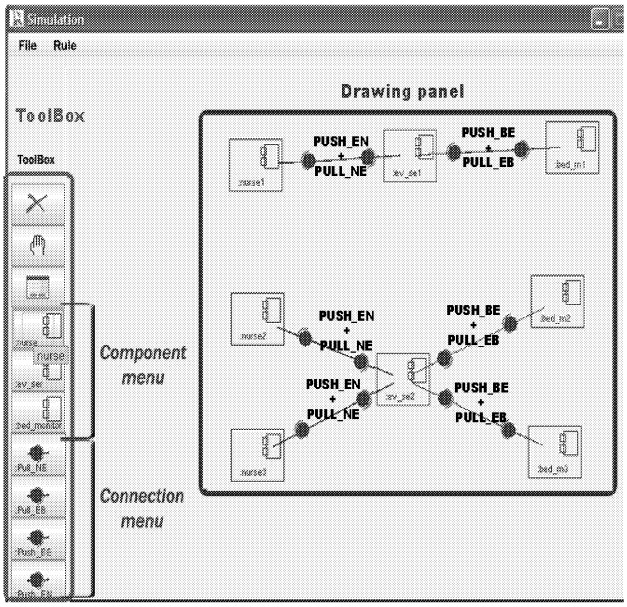


Figure 1: The simulator GUI

We group in figure 2 the different operating scenarios of our simulation environment. Its use proceeds according to following phases :

1. The start up of the simulator.
2. The user commands to the simulator to read the specification (with the Z notation) to simulate.
3. The simulator parses the specification in order to extract the architectural style (the types of components and connections, as well as the constraints) and to list the reconfiguration operations.
4. The simulator prepares the toolbox (for each component or connection type corresponds a button).
5. The simulation of reconfigurations: the user press on a button in the toolbox (which represents a component type or a connection type) and clicks on the drawing panel in order to represent graphically the component chosen according to the UML 2.0 notation.
6. The simulator checks the compatibility of each reconfiguration with the architectural

style. If it generates an architecture instance which respects the style, it executes the reconfiguration, otherwise it breaks it.

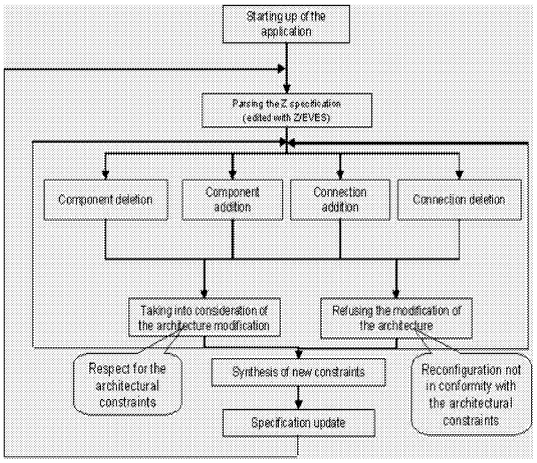


Figure 2: Functional scenario

7. If during simulation the user detects a case which does not comply with the requirements, then we are brought to the initial Z specification. To take into account of the new synthesized property, we use the command *Update* of the menu *Rule*. Thus we put the new changes into practice.

In the following, we present the use-cases and the class diagram designing the simulator. This latter has been implemented in Java.

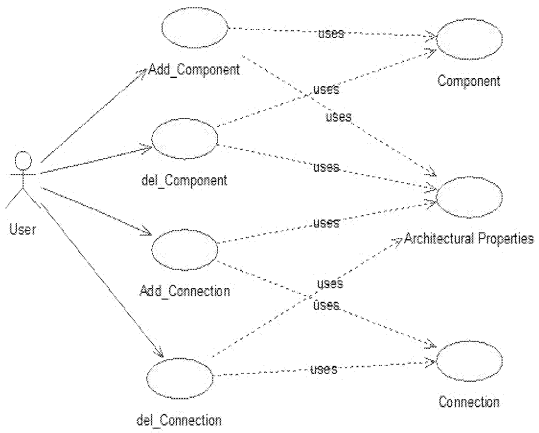


Figure 3: Use-cases Diagram (UML)

The use-cases diagram (figure 3) is the projection of the actions to accomplish by the user, namely, adding/removing components and connections. The application of these actions proceeds while respecting the architectural properties.

The fact that our parsing of the Z/EVES file is based on tags, makes our system ready to support any architectural style and any operation of reconfiguration. This also returns our tool generic and ready to deal with all architecture specifications edited with Z/EVES according to the approach presented in section 3.

In the following section, we describe the coupling of the simulator with the Z/EVES specification tool.

5 Coupling the simulator with Z/EVES

The first step of our simulation process is the specification of the architecture as described in section 3. The second step is the visual simulation (based en the Z specification) with the UML2.0 notation. We use Z/EVES to help users for editing and checking specification. Also, this tool provide a proof support with less user interaction. With the theorem prover of Z/EVES, we can prove the specified systems. Therefore, we check the consistence of the architectural style. During the second step, users can conceive the architecture with UML 2.0 notation. Thus, we can verify and validate the conformity of the conceived diagram with the requirements specification. Moreover, we provide the possibility of modifying/updating the Z specification at any time without suspending the current simulation process (cf figure 4). We terminate the modification when we obtain a complete specification.

To take into account of the architectural style during graphic simulation, our environment proceeds by identifying, in the Z specification (in its text format), of the component and connection types, and the architectural constraints.

Our simulator examines and translates the architectural properties into a set of constraints to be respected by the user during the simulation process. If the latter does not obey the constraints, the simulator prohibits it while showing an error message. For example, when we begin with configuration containing three events ser-

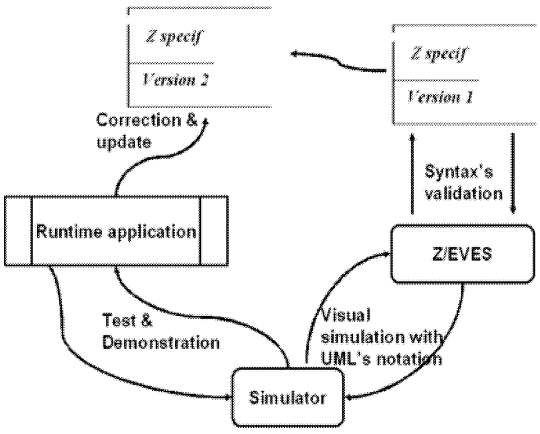


Figure 4: The simulation process

vices and we want to add the fourth, then this reconfiguration creates an inconsistent state ($|\#$ ES < 3 in schema 2). Therefore, it's ignored. Also, we cannot bind two components as long as there is no connection specified in the architectural style between them. For example, it is impossible to bind a nurse with a bed monitor because no relation between them is defined in the specification. Likewise, we cannot connect two components with the same type (two events services, for example). Additionally, we can update the specification whenever we want to satisfy the user's need. For example : Due to the small number of patients and if the clinic manager wants to reduce the expenses, he must limit the nurses's number to twenty. To allow the taking into account of the new rule, we add the following line to the architectural properties of PMS : $|\#$ CN < 20 and we reload the new specification which forbids any configuration to simulate with a nurses number superior to twenty.

6 Conclusion

We presented in this paper a simulator for controlling, by visual simulation, the model of software architectures. Our purpose is to provide to the designers a visual tool allowing the formal specification of correct and complete dynamic architectures. The graphic representation, which our simulator generates, establishes an automatic translation of a formal specification towards the visual notation UML 2.0. The interest of this environment is to enable designers, who are not familiar with formal techniques, to better understand the architectural properties of the system formally specified with Z. The graphic visualization of an erroneous configuration helps to identify the specification defects and to rectify them. This lead to have a consistence and a complete specification which will be conform with the requirements engineering.
In perspective, we intend to take into account the behaviour of components during the simulation process. For that goal, we envisage the integration of a formal language which supports the description of processes, such as process algebra, Petri nets or temporal logic.

References

[1] Ibepace. available at http://www.ibepace.com/,.

[2] Objectgeode. available at http://www.telelogic.com/products/additional/objectgeode.

[3] Objectime. available at http://www.objectime.com/,.

[4] R. Allen, R. Douence, and D. Garlan. *Specifying and analyzing dynamic software architectures*. Lecture Notes in Computer Science. 1998.

[5] C. Canal, E. Pimentel, and J. M. Troya. Specification and refinement of dynamic software architectures. In *In Software Architecture*, pages 107–126. Kluwer Academic Publishing, February 1999.

[6] Object Management Group. Uml superstructure 2.0 - draft adopted specification. 2003.

[7] M. A. Hewitt, Colin O'Halloran, and Chris T. Sennett. Experiences with piza, an animator for z. In *ZUM '97: Proceedings of the 10th International Conference of Z Users on The Z Formal Specification Notation*, pages 37–51. Springer-Verlag, 1997.

[8] Xiaoping Jia. An approach to animating z specifications. In *COMPSAC '95: Proceedings of the 19th International Computer Software and Applications Confer-*

ence, page 108. IEEE Computer Society, 1995.

[9] I. Loulou, A. Hadj Kacem, M. Jmaiel, and K. Drira. Toward a unified graph-based framework for dynamic component-based architectures description in z. In *ACS/IEEE International Conference on Pervasive Services ICPS'04*, 2004.

[10] D. C. Luckham, J. L. Kenney, L. M. Augustin, J. Vera, D. Bryan, and W. Mann. Specification and analysis of system architecture using rapide. In *IEEE Trans. on Software Engineering*, volume 21, pages 336–355, 1995.

[11] J. Magee and J. Kramer. Dynamic structure in software architectures. In *the ACM SIGSOFT Symp. on Foundations of Software Engineering*, pages 3–14, 1996.

[12] D. Le Métayer. Describing software architecture styles using graph grammars. *IEEE Transactions on Software Engineering*, 24(7):521–553, July 1998.

[13] M.Utting. Jaza user manual and tutorial. http://www.cs.waikato.ac.nz/ marku/jaza/.

[14] ORA. Z/eves. http://www.ora.on.ca/Z-eves.

[15] J.M. Spivey. *The Z Notation : a Reference Manual, Series in Computer Science*. Prentice-Hall, 1992.

[16] J. Vera, L. Perrochon, and D. C. Luckham. Event-based execution architectures for dynamic software systems. In *the Working IFIP Conf. on Software Architecture*, pages 22–24, 1999.

[17] Björn Victor and Faron Moller. The Mobility Workbench — a tool for the π-calculus. In David Dill, editor, *CAV'94: Computer Aided Verification*, volume 818 of *Lecture Notes in Computer Science*, pages 428–440. Springer-Verlag, 1994.

[18] I. Warren and I. Sommerville. Dynamic configuration abstraction. In W. Schäfer and P. Botella, editors, *the First European Software Engineering Conférence*. Springer-Verlag, 1995.

Path following for underactuated marine craft using Line Of Sight algorithm

J.GHOMMEM[1,2], F.MNIF[1] and N.DERBEL[1]

[1]*Intelligent Control Design and Optimization of Complex Systems Unit, ENIS, Tunisia*
[2]*Laboratoire de Vision et de Robotique, Orléans, France*
Email: jawhar.ghommam@ensi-bourges.fr, mnif@squ.edu.om and n.derbel@enis.rnu.tn

***Abstract*— This work provides algorithms for underactuated ship path guidance using nonlinear control theory. Path following is achieved by a geometric assignment based on a line-of-sight projection algorithm for minimization of the cross-track error to the path. The control laws in surge and yaw are derived using backstepping. This results in a dynamic feedback controller where the dynamics of the uncontrolled sway mode enters the yaw control law. Uniform Global Asymptotic Stability (UGAS) is proven for the tracking error dynamics in surge and yaw while the controller dynamics is bounded. To show the effectiveness of the controller and guidance systems, we simulate our algorithm using Simulink on a marine craft model.**

***Index Terms*— Nonlinear control, underactuated marine craft, backstepping control, Line of Sight, Global uniform asymptotic convergence.**

I. INTRODUCTION

For floating rigs and supply vessels, trajectory tracking in surge, sway, and yaw (3 DOF) is easily achieved since independent control forces and moments are simultaneously available in all degrees of freedom. For slow speed, this is referred to as dynamic positioning (DP) where the ship is controlled by means of tunnel thrusters, azimuths, and main propellers; see (*Fossen 2002*). Conventional ships, on the other hand, are usually equipped with one or two main propellers for forward speed control and rudders for turning control. The minimum configuration for way-point tracking control is one main propeller and a single rudder. This means that only two controls are available, thus rendering the ship underactuated for the task of 3 DOF tracking control. Recently, underactuated tracking control in 3 DOF has been addressed by (*Pettersen and Nijmeijer 1999, 2001*), (*Jiang and Nijmeijer 1999*), (*Sira-Ramirez 1999*), (*Jiang 2002*), (*Do et al. (2002)*), and (*Lefeber et al. 2003*). These designs deals with simultaneous tracking control in all three modes (x, y, ψ) using only two controls. One of the main problems with this approach is that integral action, needed for compensation of slowly-varying disturbances due to wind, waves, and currents, can only be assigned to two modes (surge and yaw); (*Pettersen and Fossen 2000*). Consequently, robustness to environmental disturbances is one limiting factor for these methods. In addition, requirements for a persistently exciting reference yaw velocity results in unrealistic topological restrictions on which type of paths that can be tracked by these controllers (*Lefeber et al. 2003*), designed by reducing the output space from 3 DOF position and heading to 2 DOF heading and surge (*Healey and Marco 1992*). In its simplest form, this involves the use of a classical autopilot system where the commanded yaw angle ψ_d is generated such that the cross-track error is minimized. This can be done in a multivariable controller, for instance H^∞ or LQG, or by including an additional tracking error control-loop in the autopilot; see (*Holzhter and Schultze 1996*), and (*Holzhter 1997*). A path following control system is usually designed such that the ship moves forward with reference speed u_d at the same time as the cross-track error to the path is minimized. As a result, ψ_d and u_d are tracked using only two controls. The desired path can be generated using a route management system or by specifying way-points (*Fossen 2002*). If weather data are available, the optimal route can be generated such that the effects of wind and water resistance are minimized.
In this paper, we suggest to use an intuitive comprehension of the shipbehavior and the action of a helmsman in order to track a stright line. We will consider the problem of ship maneuvering design involving a Line Of Sight guidance system and a nonlinear feedback tracking controller. The concept of the LOS angle can be explained by utilizing the terms LOS vector and LOS position. The LOS vector is a vector pointing from the current position of the ship towards a position along the desired geometrical path, the so-called LOS position. This position can be calculated in many different ways. The angle which the LOS vector makes with the x-axis in the NED reference frame is defined to be the LOS angle, i.e. the ship should simply point towards the LOS position.

Having introduced the aim of this work, we will now outline the reminder of this paper. In section 2, kinematic and dynamic of a three DOF marine craft are presented. In section 3, the problem statement for the maneuvering of a marine craft using the LOS algorithm is presented. In section 4 the LOS controller algorithm is presented. Simulation results illustrating the convergence of the marine craft to the desired trajectory are presented in section 5 and some conclusions are given in section 6

II. SHIP'S DYNAMIC MODEL

A. Kinematic

For commercial ships it is common to reduce the general 6 DOF model to surge, sway and yaw by neglecting the heave,

roll and pitch modes. The state vector $\eta \in \mathbf{SE(2)}$ is defined

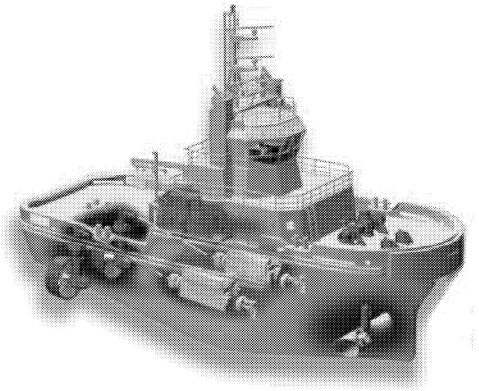


Fig. 1. Drawing of a tugboat with two main azimuth thrusters aft and one tunnel thruster in the bow. Courtesy of Rolls-Royce Marine [22].

by :

$$\eta = [x, y, \psi]^T \tag{1}$$

where $(x, y) \in R^2$ is the position of the ship given in an inertial frame, and $\psi \in [0, 2\pi[$ is the heading angle of the ship relative to the geographic North.
The kinematics is written as:

$$\dot{\eta} = \mathbf{R}(\psi)\nu \tag{2}$$

where $\mathbf{R}$ is the rotation matrix in yaw defined as

$$\mathbf{R}(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

with the property $\mathbf{R^T} = \mathbf{R^{-1}}$. The time derivative of $\mathbf{R}$ is given by

$$\dot{\mathbf{R}}(\psi, r) = \begin{bmatrix} -\sin(\psi) & -\cos(\psi) & 0 \\ \cos(\psi) & sin(\psi) & 0 \\ 0 & 0 & 0 \end{bmatrix} r \tag{4}$$

where r is the angular velocity in yaw. Let $\nu \in \mathbb{R}^3$ be the linear body fixed velocities in surge, sway and yaw

$$\nu = [u, v, r]^T \tag{5}$$

Here u is the forward velocity (surge) and v is the transverse velocity (sway).

B. Ship's Dynamic

The nonlinear ship model used in this paper is based on the work of Fossen (1994). The model describes the motion of a ship in surge, sway and yaw. Denote the control forces in surge, sway and the yaw moment by τ. The nonlinear ship model is given by:

$$\mathbf{M}(\nu)\dot{\nu} + \mathbf{n}(\nu) = \tau \tag{6}$$

where the Coriolis and centripetal matrix $\mathbf{C}$ and damping matrix $\mathbf{D}$ are collected into a vector $\mathbf{n}$ given by

$$\mathbf{n}(\nu) = \mathbf{C}(\nu)\nu + \mathbf{D}(\nu)\nu \tag{7}$$

the expressions of $\mathbf{M}, \mathbf{C}(\nu), \mathbf{D}(\nu)$ are respectively given by The inertia matrix

$$\mathbf{M} = \mathbf{diag}(m_{11}, m_{22,} m_{33})$$

$$\mathbf{C}(\nu) = \begin{bmatrix} 0 & 0 & -m_{22}v \\ 0 & 0 & m_{11}u \\ m_{22}v & -m_{11}u & 0 \end{bmatrix}$$

$$\mathbf{D}(\nu) = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & d_{23} \\ 0 & d_{32} & d_{33} \end{bmatrix}$$

$\mathbf{M}$ has the following proprieties :

- $\mathbf{M} = \mathbf{M^T} > 0$
- $\dot{\mathbf{M}} - 2\mathbf{C}$ *is skew symmetric matrix*

C. Environmental Disturbances

Constant or slowly varying environmental disturbances due to ocean currents, 2nd-order wave drift and wind can be compensated by using adaptive backstepping. The environmental disturbences are usually modelled as a constant bias force with respect to the inertial reference frame, that is

$$\dot{\mathbf{b}} = 0$$

Hence, the body fixed environmental force $\mathbf{R^T}(\psi)\mathbf{b}$ can be added to the ship dynamics according to:

$$\mathbf{M}(\nu)\dot{\nu} + \mathbf{n}(\nu) = \tau + \mathbf{R^T}(\psi)\mathbf{b} \tag{8}$$

The overall marine craft's dynamic is then described by both the kinematic and dynamic model of the ship as follow

$$\dot{\eta} = \mathbf{R}(\psi)\nu \tag{9}$$

$$\mathbf{M}(\nu)\dot{\nu} + \mathbf{n}(\nu) = \tau + \mathbf{R^T}(\psi)\mathbf{b} \tag{10}$$

Notice that in our application we assume that the tunel thruster located in the bow (fig.1) looses its capability to steer the motion of the ship on the yaw direction when the speed of the ship increases significantly during maneuvering. Hence we consider through out this paper that the control input $\tau = [\tau_1, 0, \tau_3]$.

We will consider the dynamic (9)-(10) to derive a control law for the maneuvering problem that will be discussed in the following section.

III. THE MANEUVERING PROBLEM FOR UNDERACTUATED MARINE CRAFT

A. Problem statement (Skjetne. 2005)

For a system output $y \in \mathbb{R}^m$, the desired path in all points is represented by the set:

$$\mathcal{P} = \{y \in \mathbb{R}^m / \exists \theta \in \mathbb{R} \; s.t. y = y_d(\theta)\} \tag{11}$$

where y_d is continuously parameterized by θ. Given the desired path (11) and a dynamic assignment, the **Maneuvering Problem** is comprised of the two tasks:

1) ***Geometric Task*** *For any continuous function $\theta(t)$; force the output y to converge to the desired path $y_d(\theta)$;*

$$\lim_{t\to\infty} |y(t) - y_d(\theta(t))| = 0$$

2) ***Dynamic Task*** *Satisfy one or more of the following assignments:*

2.1 ***Time Assignment:*** *Force the path variable θ to converge to a desired time signal $v_t(t)$,*

$$\lim_{t\to\infty} |\theta - v_t(t)| = 0$$

2.2 ***Speed Assignment:*** *Force the path speed $\dot{\theta}$ to converge to a desired speed $v_s(\theta, t)$*

$$\lim_{t\to\infty} |\dot{\theta} - v_s(\theta, t)| = 0$$

2.3 ***Acceleration Assignment:*** *Force the path acceleration $\ddot{\theta}$ to converge to a desired acceleration $v_a(\dot{\theta}, \theta, t)$*

$$\lim_{t\to\infty} |\ddot{\theta} - v_a(\dot{\theta}, \theta, t)| = 0$$

B. Path generation for a 3-DOF underactuate ship

To solve this as a maneuvering problem we let the desired path be

$$\eta_d(\theta) = [x_d(\theta), y_d(\theta), \psi(\theta)]^T$$

where $p_d(\theta) = [x_d(\theta), y_d(\theta)]^T$ is a curve generated by a table of way points. One choice for the desired heading $\psi_d(\theta)$ is a constant reference $\psi_d(\theta) = \psi_{ref}$, or it can be chosen as the angle of the tangent vector along the path (Breivik 2003)

$$\psi_d(\theta) = \arctan 2(y_d^\theta(\theta), x_d^\theta(\theta)) \tag{12}$$

For the speed assignment $v_s(\theta, t)$ we let a desired path speed (in m/s) be a commanded input speed $u_d(t)$. Since the identity

$$\begin{aligned} |\dot{p}_d(\theta(t))| &= \sqrt{(x_d^2 + y_d^2)\dot{\theta}^2} \\ &= \sqrt{x_d^2 + y_d^2}|v_s| = |u_d| \\ v_s &= \frac{u_d}{\sqrt{x_d^2 + y_d^2}} \end{aligned}$$

hence if $u_d = 0$ the ship will stop on the path. However if $u_d > 0$ the ship will move in positive direction while it will move in a negative direction if $u_d < 0$.

The main idea for the path generation is to make the ship's heading to converge to a point p_{los} located somewhere along the straight-line segment connecting the previous p_{k-1} and current p_k way-points (fig.2).

The geometric task for maneuvering may be reformulated as follows: *Forcing the ship position p to converge to a desired path by forcing the heading angle ψ to converge to the LOS angle ψ_{los}. According to (12) ψ_{los} is given by*

$$\begin{aligned} \psi_{los} &= atan2(y_{los} - y, x_{los} - x) = f(p_{los}) && (13) \\ p_{los} &= \min_{dist} \{p_k, (p_{k-1}, p_k) \cap \mathcal{C}_{p,R_p}\} && (14) \end{aligned}$$

which means that:

$$\lim_{t\to\infty} |\psi(t) - \psi_{los}(t)| = 0$$

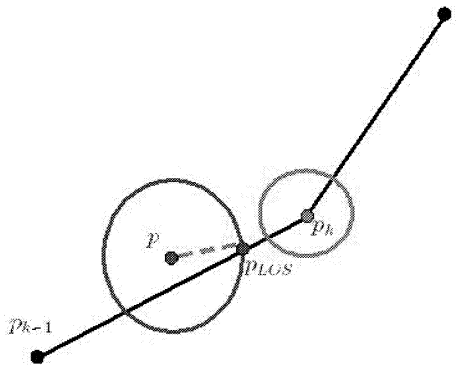


Fig. 2. The Line-of-Sight principle.

The purpose of the controller will be to find τ such that ψ converges to ψ_{los} in a finite time. This will be the aim of the next section.

IV. BACKSTEPPING CONTROL DESIGN

The backstepping design is performed in two steps. First define the error variable $z_1 \in \mathbb{R}$ and $z_2 \in \mathbb{R}^3$ according to:

$$\begin{aligned} z_1 &= \psi - \psi_d = \Pi^T \eta - \psi_d && (15) \\ z_2 &= \nu - \alpha && (16) \end{aligned}$$

where $\alpha \in \mathbb{R}^3$ is a vector of stabilizing function to be specified later and $\Pi^T = [0\ 0\ 1]$.

a) ***Step I:*** We start by defining the first control Lyapunov function to be

$$\mathbf{V_1} = \frac{1}{2} z_1^2 \tag{17}$$

Its time derivative yields:

$$\begin{aligned} \dot{\mathbf{V}}_\mathbf{1} = z_1 \dot{z}_1 &= z_1(\Pi^T \dot{\eta} - \dot{\psi}_d) \\ &= z_1(\Pi^T \mathbf{R}\nu - \dot{\psi}_d) \\ &= z_1(\Pi^T \nu - \dot{\psi}_d) \\ &= z_1(\Pi^T (z_2 + \alpha) - \dot{\psi}_d) \\ &= z_1 \Pi^T z_2 + z_1(\Pi^T \alpha - \dot{\psi}_d) && (18) \end{aligned}$$

Since $\Pi^T \alpha = \alpha_3$ is a scalar, motivates to chose α_3 as a stabilizing function to be:

$$\alpha_3 = -k_1 z_1 + \dot{\psi}_d \tag{19}$$

the expression (19) becomes

$$\dot{\mathbf{V}}_\mathbf{1} = -k_1 z_1^2 + z_1 \Pi^T z_2 \tag{20}$$

this will end the first step.

b) ***Step II:*** In the second step we define an augmented control Lyapunov to the previous one, to obtain:

$$\mathbf{V}_2 = \mathbf{V}_1 + \frac{1}{2} z_2^T \mathbf{M} z_2 + \frac{1}{2} \tilde{\mathbf{b}}^T \mathbf{\Gamma^{-1}} \tilde{\mathbf{b}} \tag{21}$$

where $\tilde{\mathbf{b}}$ is the adaptation error defined as $\tilde{b} = \hat{b} - b$. The term $\hat{\mathbf{b}}$ expresses the estimate of the environmental disturbance vector b. The matrix $\mathbf{\Gamma} = \mathbf{\Gamma}^T > 0$ is the adaptation gain matrix.

For a practical purpose, we assume that $\dot{\mathbf{b}} = \mathbf{0}$ since we have supposed that the $\mathbf{b}$ is slowly time varying.

Differentiating $\mathbf{V_2}$ along the trajectories of z_1 and z_2 and $\tilde{\mathbf{b}}$, yields

$$\dot{\mathbf{V}}_\mathbf{2} = -k_1 z_1^2 + z_1 \Pi^T z_2 + \frac{1}{2} z_2^T \dot{\mathbf{M}} z_2 + z_2^T \mathbf{M} \dot{z}_2 + \tilde{\mathbf{b}}^\mathbf{T} \mathbf{\Gamma^{-1}} \dot{\tilde{\mathbf{b}}} \quad (22)$$

From (10) and (16) we can write that

$$\begin{aligned} \mathbf{M}\dot{z}_2 &= \mathbf{M}(\dot{v} - \dot{\alpha}) \\ &= \tau + \mathbf{R^T b} - \mathbf{n}(\nu) - \mathbf{M}\dot{\alpha} \end{aligned} \quad (23)$$

which yields :

$$\begin{aligned} \dot{\mathbf{V}}_\mathbf{2} &= -k_1 z_1^2 + \frac{1}{2} z_2^T \dot{\mathbf{M}} z_2 \\ &+ z_2^T (\Pi z_1 + \tau + \mathbf{R^T b} - \mathbf{n}(\nu) - \mathbf{M}\dot{\alpha}) \\ &+ \tilde{\mathbf{b}}^\mathbf{T} \mathbf{\Gamma^{-1}} \dot{\hat{\mathbf{b}}} \end{aligned} \quad (24)$$

From equations (7) and (16), we can write that

$$\mathbf{n}(\nu) = \mathbf{C}(\nu)\alpha + \mathbf{C}(\nu) z_2 + \mathbf{D}(\nu)\alpha + \mathbf{D}(\nu) z_2 \quad (25)$$

Substituting (25) in (24), we obtain

$$\begin{aligned} \dot{\mathbf{V}}_\mathbf{2} &= -k_1 z_1^2 + \frac{1}{2} z_2^T (\dot{\mathbf{M}} - 2\mathbf{C}) z_2 - z_2^T \mathbf{D} z_2 \\ &+ z_2^T (\Pi z_1 + \tau + \mathbf{R^T \hat{b}} - \mathbf{C}\alpha - \mathbf{D}\alpha - \mathbf{M}\dot{\alpha}) \\ &+ \tilde{\mathbf{b}}^\mathbf{T} \mathbf{\Gamma^{-1}} (\dot{\hat{\mathbf{b}}} - \mathbf{\Gamma R} z_2) \end{aligned} \quad (26)$$

Since the matrix $(\dot{\mathbf{M}} - 2\mathbf{C})$ is skew symmetric then $z_2^T (\dot{\mathbf{M}} - 2\mathbf{C}) z_2 = 0$. Assuming that the ship is fully actuated, i.e $\tau = [\tau_1, \tau_2, \tau_3]$ then it will be possible to assign the control law τ the following expression

$$\tau = -\mathbf{K}_2 z_2 - \Pi z_1 - \mathbf{R^T \hat{b}} + \mathbf{C}\alpha + \mathbf{D}\alpha + \mathbf{M}\dot{\alpha} \quad (27)$$

In order to cancel the adaptation error from the equation (27), we choose the updated law $\dot{\hat{\mathbf{b}}}$ to be:

$$\dot{\hat{\mathbf{b}}} = \mathbf{\Gamma R} z_2 \quad (28)$$

This yields :

$$\dot{\mathbf{V}}_\mathbf{2} = -k_1 z_1^2 - z_2^T (\mathbf{D} + \mathbf{K_2}) z_2 \leq 0 \quad (29)$$

Currently, α_3 is the only stabilizing function which has been defined. Since we want the surge speed of the vessel to track a given speed assignment, we choose $\alpha_1 = u_d$. It is also desirable that the sway velocity of the vessel is kept identically at zero, such that the vessel traverses the path with its heading aligned with the path tangential vector. Consequently, we assign $\alpha_2 = 0$, which is feasible since the sway direction of the vessel is actuated. Hence, we obtain:

$$\alpha = [u_d, 0, -k_1 z_1 + \dot{\psi}_d]$$

Error dynamics:
Let's define $\chi = [z_1, z_2]^T$. The error dynamic in terms of the vector χ of the resulting system becomes :

$$\dot{\chi} = \begin{pmatrix} -k_1 & \Pi^T \\ -\Pi & \Xi \end{pmatrix} \chi + \begin{pmatrix} 0 \\ \Theta \end{pmatrix} \tilde{\mathbf{b}} \quad (30)$$

$$\dot{\tilde{\mathbf{b}}} = \mathbf{\Gamma R} z_2 \quad (31)$$

where $\Xi = -\mathbf{M^{-1}}(\mathbf{C} + D + \mathbf{K_2})$ and $\Theta = -\mathbf{M^{-1}R^T}$.

Theorem 1 (Loria et al, 1999): Consider the nonlinear system :

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \theta, t)$$

where $x \in \mathbb{R}^n$; $u \in \mathbb{R}^n$, and $\theta \in \mathbb{R}^p (p \leq n)$ is a constant unknown parameter vector. Furthermore, assume that there exists an adaptive control law :

$$\mathbf{u} = \mathbf{u}(x, x_d, \hat{\theta}) \quad (32)$$

$$\dot{\hat{\theta}} = \phi(x, x_d) \quad (33)$$

such that the error dynamics can be written as:

$$\dot{\chi} = h(\chi, t) + G(\chi, t)\tilde{\theta} \quad (34)$$

$$\dot{\tilde{\theta}} = -PG(\chi, t)^T \left(\frac{\partial \mathbf{W}(\chi, t)}{\partial \chi}\right), \quad P = P^T > 0 \quad (35)$$

where $\mathbf{W}(\chi, t)$ is a suitable $\mathcal{C}^\infty$ function and $\tilde{\theta} = \hat{\theta} - \theta$ is the parameter estimation error. The origin $(\chi, \tilde{\theta})$ is uniformly globally asymptotically stable if there exist a continuous, non-decreasing function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ such that:

$$max \left\{ ||h(\chi, t)||, ||\frac{\partial \mathbf{W}(\chi, t)}{\partial \chi}|| \right\} \leq \rho(||\chi||)||\chi|| \quad (36)$$

and there exist class $\mathcal{K}_\infty$ functions β_1 and β_2 and a strictly positive real number $c > 0$ such that $\mathbf{W}(\chi, t)$ stratifies

$$\beta_1(\|\chi\|) \leq \mathbf{W}(\chi, t) \leq \beta_2(\|\chi\|) \quad (37)$$

$$\frac{\partial \mathbf{W}(\chi,t)}{\partial t} + \frac{\partial \mathbf{W}(\chi,t)}{\partial \chi} \mathbf{h}(\chi,t) \leq -c\|\chi\|^2 \quad (38)$$

By direct application of this theorem to system (30)-(31), since $||h(\chi, t)|| = \left\| \begin{pmatrix} -k_1 & \Pi^T \\ -\Pi & \Xi \end{pmatrix} \right\|$ and $||\frac{\partial \mathbf{W}(\chi, \mathbf{t})}{\partial \chi}|| = ||\chi^T \overline{\mathbf{M}}||$ then there exists a continuous and non decreasing function satisfying (37). Moreover we have $\mathbf{W}(\chi, \mathbf{t}) = \frac{1}{2}\chi^T \overline{\mathbf{M}} \chi$, then there exist $\beta_1 = \frac{1}{4}\mathbf{W}(\chi, \mathbf{t})$ and $\beta_2 = 2\mathbf{W}(\chi, \mathbf{t})$ such that condition (38) is satisfied. Here we notice that

$$\overline{\mathbf{M}} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{M} \end{pmatrix}$$

Now the time derivative of $\mathbf{W}(\chi, \mathbf{t})$ is

$$\frac{\partial \mathbf{W}}{\partial t} = \chi^T (\frac{1}{2}\dot{\overline{\mathbf{M}}} - \overline{\mathbf{C}} - \overline{\mathbf{D}})\chi = -\chi^T \overline{\mathbf{D}} \chi$$

and

$$\frac{\partial \mathbf{W}}{\partial \chi} h(\chi, t) = \chi^T \begin{pmatrix} -k_1 & 0 \\ 0 & \mathbf{M\Xi} \end{pmatrix} \chi < 0$$

.
Consequently:

$$\frac{\partial \mathbf{W}(\chi,t)}{\partial t} + \frac{\partial \mathbf{W}(\chi,t)}{\partial \chi} \mathbf{h}(\chi,t) =$$
$$\chi^T \begin{pmatrix} -k_1 & 0 \\ 0 & \mathbf{M\Xi} \end{pmatrix} \chi - \chi^T \overline{\mathbf{D}} \chi < -c||\chi||^2$$

The fulfillment of the theorem renders the origin $(\chi, \tilde{b})$ UGAS, and the angle $\psi \to \psi_d$

In fact since we considered that the ship is underactuated in terms of the number of control inputs i.e $\tau_2 = 0$, hence a second analysis step must be performed:
If we rewrite the control law (27), with the assumption that $\tau_2 = 0$, we may write that :

$$\begin{bmatrix} \tau_1 \\ 0 \\ \tau_2 \end{bmatrix} = -\mathbf{K}_2 z_2 - \Pi z_1 - \mathbf{R}^{\mathbf{T}}\hat{\mathbf{b}} + \mathbf{C}\alpha + \mathbf{D}\alpha + \mathbf{M}\dot{\alpha} \quad (39)$$

notice from (39) we can only prescribe values for τ_1 and τ_3. The remaining equation ($\tau_2 = 0$) in (39) results in a dynamic equality constraint given by (40)

$$\begin{aligned} \mathbf{\Pi_2} &= [0\ 1\ 0]^T \\ 0 &= \mathbf{\Pi}_2^T(\mathbf{M}\dot{\alpha} + \mathbf{N}\alpha - \mathbf{K}_2 z_2 - \Pi z_1 - \mathbf{R}^T\hat{\mathbf{b}}) \\ m_{22}\dot{\alpha}_2 &= -(d_{22} + \mathbf{n}_{22})\alpha_2 + \mathbf{k}_{22} v \\ &- [d_{22}\alpha_3 + (\hat{b}_1 \sin(\psi) - \hat{b}_2 \cos(\psi))] \end{aligned}$$

Since $\nu = z_2 + \alpha$, hence $v = \mathbf{\Pi}_2^T \nu = \mathbf{\Pi}_2^T z_2 + \alpha_2$, then

$$\begin{aligned} m_{22}\dot{\alpha}_2 &= -(d_{22} + \mathbf{n}_{22} - \mathbf{k}_{22})\alpha_2 \\ &- [d_{22}(-k_1 z_1 + \dot{\psi}_d) + (\hat{b}_1 \sin(\psi) - \hat{b}_2 \cos(\psi)) \\ &- \mathbf{\Pi}_2^T z_2] \\ &= -(d_{22} + \mathbf{n}_{22} - \mathbf{k}_{22})\alpha_2 + \gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d) \quad (40) \end{aligned}$$

Choosing $\mathbf{k}_{22} < d_{22} + \mathbf{n}_{22}$ implies that (41) without the perturbed term γ i.e ($\gamma = 0$) is a stable differential equation. Moreover, according to (27) and using Barbalat's theorem (*Slotine 1991*), the couple $(z_1, z_2) \in \mathcal{L}_\infty$ then $\hat{\mathbf{b}} \in \mathcal{L}_\infty$ and $\gamma \in \mathcal{L}_\infty$. This implies that α_2 is *Input to State Stable* (**ISS**) from γ to α_2 (*Sontag 1995*). This is seen by applying for instance the following Lyapunov function $V_0 = \alpha_2^2$, we have

$$\begin{aligned} \dot{V}_0 &= -(d_{22} + \mathbf{n}_{22} - \mathbf{k}_{22})\frac{dV_0}{d\alpha_2}\alpha_2 + 2\alpha_2\gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d) \\ &\leq 2\sqrt{V_0}|\gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d)| \end{aligned}$$

for $\alpha_2 \neq 0$, we can afford to divide the previous inequality by $2\sqrt{V_0}$, then integrating both sides over the interval $[0, T)$, one obtains

$$\begin{aligned} \sqrt{V_0(\alpha_2(T)} - \sqrt{V_0(\alpha_2(0)} &\leq \int_0^T |\gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d)|dt \\ &\leq \int_0^\infty |\gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d)|dt \\ &\leq l(\epsilon) \end{aligned}$$

where ϵ and $l(\epsilon)$ are positive constants, then for all $T > 0$

$$V_0(\alpha_2(\alpha(T))) \leq (l(\epsilon) + \sqrt{V_0(\alpha_2(0))})^2$$

which implies boundedness of α_2. Since γ is vanishing, after some finite time $T_\xi > 0$, α_2 enters a closed loop neighborhood $\Omega_\xi = \{\alpha_2 : |\alpha_2| \leq \xi\}$ of $\alpha_2 = 0$ This establishes global asymptotic and stability of $\alpha_2 = 0$ for the α_2-subsystem of (41).
The main result is summarized by the following Theorem.

Theorem 2: The maneuvering problem for the 3 DOF underactuated vessel model (9)(10) *is solved using the control laws (39):*

$$\begin{bmatrix} \tau_1 \\ 0 \\ \tau_2 \end{bmatrix} = -\mathbf{K}_2 z_2 - \Pi z_1 - \mathbf{R}^{\mathbf{T}}\hat{\mathbf{b}} + \mathbf{C}\alpha + \mathbf{D}\alpha + \mathbf{M}\dot{\alpha}$$

while the constrained equation is found by numerical integration of:

$$m_{22}\dot{\alpha}_2 = -(d_{22} + \mathbf{n}_{22} - \mathbf{k}_{22})\alpha_2 + \gamma(z_1, z_2, \hat{\mathbf{b}}, \psi_d)$$

This results in a UGAS equilibrium point $(z_1, z_2) = 0$*, while* $\alpha_2 \in \mathcal{L}_\infty$ *satisfies:*

$$\lim_{t\to\infty} |\alpha_2 - v(t)| = 0$$

In order to unify the control law for both fully actuated and underactuated ship, we propose the following formula for a transition between the two modes:
Let's define a weighting variable $\sigma \in [0\ 1]$ with the following properties :

$$\sigma = \begin{cases} 1, & \text{if the ship is fully actuated;} \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

then the control law will be as follows:

$$\mathbf{u} = \sigma\tau_{fa} + (1 - \sigma)\tau_{ua} \quad (42)$$

Like-wise, we define a sway stabilizing function as

$$\alpha_{\mathbf{2n}} = \sigma\alpha_{fa} + (1 - \sigma)\alpha_{ua} \quad (43)$$

where τ_{fa}, α_{fa} and τ_{ua}, α_{ua} are control inputs and stabilizing functions when the ship is fully actuated and underactuated (respectively).
To ensure a smooth transition between a fully actuated and an underactuated vessel, we suggest to implement this variable as being a sigmoidal function dependent on the instantaneous vessel speed $U = \sqrt{u^2 + v^2}$. Consider the upper limit U_{ua} of U for which the ship becomes underactuated and the lower bound U_{fa} of U for which the ship still fully actuated, then the choice for the weighing variable σ could be

$$\sigma = \frac{1}{2}(1 - \mathbf{tanh}(U - \overline{U})) \quad (44)$$

where $\overline{U} = \frac{U_{fa} + U_{ua}}{2}$

V. Simulations and results

For the sake of illustration, simulations are performed with a vessel executing a straight-line while being exposed to a constant environmental force averaging $1N$. The vessel data are as follows 1,70 scale model of an offshore supply vessel, with a mass of $m = 23.8\ kg$ and a length of $L = 1.255\ m$. The Controller gains are chosen as $k_1 = 0.1$ and $\mathbf{K}_2 = 50\mathbf{I}$, while $\mathbf{\Gamma} = 70\mathbf{I}$.
Figure 3 shows the convergence of the ship's heading to the desired angle heading ψ_{los} calculated using formula (12), which forces the ship to converge to the desired position p_{los}. Figures 4 and 5 illustrate respectively the convergence of the

error variable z_2 and the dynamic of the stabilizing function α_2. The figure 6 shows the time evolution of the control inputs.

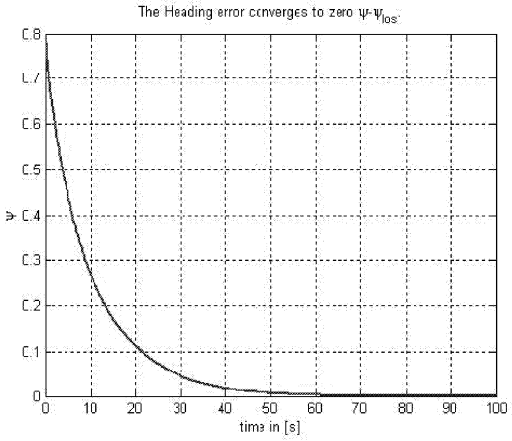


Fig. 3. Convergence of the heading angle to the desired angle LOS

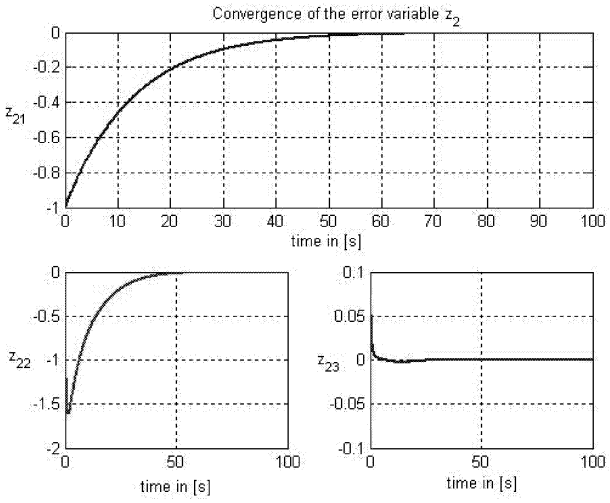


Fig. 4. Convergence of the error vector z_2

VI. Conclusion

In this paper we have presented a technique for a guidance based path following to maneuvering marine vessels. This approach guarantees position convergence to a desired straight-line path. Furthermore, a nonlinear model-based controller using backstepping technique has been designed for a fully actuated vessel such that it can comply with the guidance controls. Since the vessel is exposed to a constant environmental force, integral action is added to the controller by means of a parameter adaptation technique. The results were extended to underactuated vessels which are unactuated in the sway direction. Experiments with a model ship show the performance of the guidance and control systems.

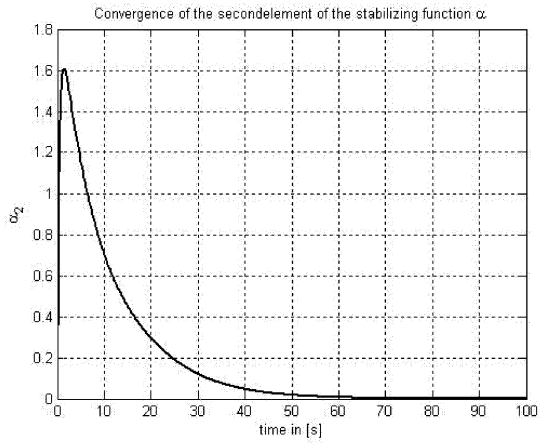


Fig. 5. Convergence of the second element of the stabilizing function α

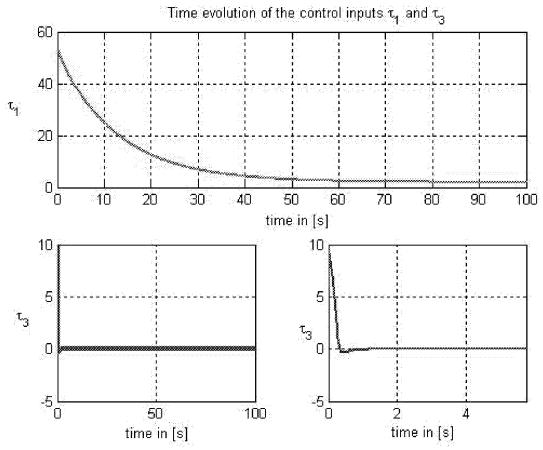


Fig. 6. Time evolution of the control inputs τ_1 and tau_3

References

[1] A.Loria, E.Panteley, and A.Teel. A new persistency-of-excitation condition for UGAS of NLTV systems: Application to stabilization of nonholonomic systems. *IN Proc*. 5th. European Contr. Conf., 1999. paper no. 500

[2] Breivik, M. (2003). Nonlinear Maneuvering Control of Underactuated Ships. MSc thesis. Dept. of Eng. Cybernetics, Norwegian University of Science and Technology.

[3] Do, K. D., Z. P. Jiang and J. Pan (2002). Underactuated Ship Global Tracking under Relaxed Conditions. IEEE Transactions on Automatic Control TAC-47(9), 1529-1535.

[4] Fossen, T. I. (1994). Guidance and Control of Ocean Vehicles. John Wiley and Sons Ltd. Norway

[5] Fossen, T. I. (2002). Marine Control Systems: Guidance,Navigation and Control of Ships, Rigs and Underwater Vehicles. Marine Cybernetics AS. Trondheim, Norway.

[6] Fossen, T. I., K. P. Lindegaard and R. Skjetne (2002). Inertia Shaping Techniques for Marine Vessels using Acceleration Feedback. In: Proceedings of the IFAC World Congress. Elsevier Science. Barcelona.

[7] Healey, A. J. and D. B. Marco (1992). Slow Speed Flight Control of Autonomous Underwater Vehicles: Experimental Results with the NPS AUV II.In: Proceedings of the 2nd International Offshore and Polar Engineering Conference (ISOPE). San Francisco, CA. pp. 523-532.

[8] Holzhter, T. (1997). LQG Approach for the High-Precision Track Control of Ships. IEE Proceedings on Control Theory and Applications 144(2), 121-127.

[9] Holzhter, T. and R. Schultze (1996). On the Experience with a High-Precision Track Controller for Commercial Ships. Control Engineering Practise CEP-4(3), 343-350.

[10] J.Ghommem, Mnif. F, Benali. A and Derbel. N:"Asymptotic Backstepping Stabilization of an Underactuated Surface Vessel".Int. Conf. Signals, Systems, Decision & information Technology, SSD05, Sousse, March 2005.
[11] Jiang, Z. P. (2002). Global Tracking Control of Underactuated Ships by Lyapunovs Direct Method. Automatica AUT-38(2), 301-309.
[12] Jiang, Z.-P. and H. Nijmeijer (1999). A Recursive Technique for Tracking Control of Nonholonomic Systems in Chained Form. IEEE Transactions on Automatic Control TAC-4(2), 265-279.
[13] Lefeber, A.A.J., K. Y. Pettersen and H. Nijmeijer (2003). Tracking Control of an Underactuated Ship. IEEE Transactions on Control Systems Technology TCST-11(1), 52-61.
[14] Pettersen, K. Y. and H. Nijmeijer (1999). Tracking Control of an Underactuated Surface Vessel. In: Proceedings of the IEEE Conference on Decision and Control. Phoenix, AZ. pp. 4561-4566.
[15] Pettersen, K. Y. and H. Nijmeijer (2001). Underactuated Ship Tracking Control. International Journal of Control IJC-74, 1435-1446.
[16] Pettersen, K. Y. and T. I. Fossen (2000). Underactuated Dynamic Positioning of a Ship - Experimental Results. IEEE Transactions on Control Systems Technology TCST-8(5), 856-863.
[17] Sira-Ramirez, H. (1999). On the Control of the Underactuated Ship: A Trajectory Planning Approach. In: IEEE Conference on Decision and Control.Phoenix, AZ.
[18] Skjetne, R., T. I. Fossen and P. V. Kokotovic (2002). Output Maneuvering for a Class of Nonlinear Systems. In: Proc. of the IFAC World Congress. Barcelona.
[19] Skjetne R., The maneuvering problem, Ph.D. dissertation, Norwegian University of Science and Technology, Trondheim, Norway, 2005.
[20] J. E. Slotine and W. Li. Applied Nonlinear Control. Prentice Hall, New Jersey, 1991.
[21] Sontag, Eduardo D. and Yuan Wang (1995). On characterizations of the input-to-state stability property. Systems and Control Letters 24, pp 351-359.
[22] http://www.rolls-royce.com/marine/

Statistical Approach for Analysis of Energy System Losses in Jordan

Nabeel Tawalbeh
Electrical Engineering Department
Isra Private University
Amman-Jordan
nabiltawalbeh@yahoo.com

Ayman Agha
Projects' Department
Jordan Phosphate Mines Co
Amman-Jordan
aay@o2.pl

KEYWORDS
Model Analysis, Regression, Forecasting, Approximation
Statistical Analysis, Energy Modeling Management

ABSTRACT

A method based on a statistical approach is presented in this paper. The regression analysis is used to evaluate, predict, and for the planning of the electrical energy data. Two regression models are used, linier and logarithmic. While both of them are suitable for electrical energy data, they are not acceptable for the analysis of the energy losses in percentage form. The level of change between the regression models is utilized to determine the relationship between the variables of interest. The proposed method is applied to the NEPCO data. Numerical examples and discussions are given to illustrate the effectiveness of the proposed method.

1. INTRODUCTION

The main aim of electrical power utilities are striving to achieve the common task of supplying reliable and affordable electrical services. Power system has to satisfy system loads and energy requirements of the consumers, taken into account the minimum possible power and energy cost, within acceptable level of availability/reliability based on the standard quality. The economic operation of the electrical power system is based on scheduling of generation, transmission and distribution system. This will provide minimum system losses.
Electrical system losses can be simply defined as a difference between the sum of electrical plants output minus the total loads consumption. Various approaches have been used to assess the loses of the electrical systems. These include various analytical methods (Chang 1968, Neagle and Samson 1956, Glamocanin 1990, Chung et al. 1994). The difference between these methods is in the way of loss calculation, simplicity and computational time. However the achieved results are the same. Traditionally the problem of energy loss computation is obtained by calculating power system losses, for a given period of time.
The trend of using statistical methods in power system is going momentum (Karakatsanis and Hatziargyriou 19994, Agha 1999). They can solve in principle rather complicated problems. The proposed method in this paper has advantages over the traditional methods that, investigate the losses behavior for a long period of time. This long-term analysis is based on historical system data for the generated, transmitted and consumed energy. It is a process, which needs the interaction of engineering issue. It could be considered as first step in the process of the system planning, evaluation and load demand forecasting. It was found that, a period of not - the analyzed period the abnormal data which could occur due to the abnormal conditions, such as deficit of generation or less consumption of energy etc.
This paper presents the application of the regression models, in evaluation and predicts the trend of system energy losses. The achieved trends can be investigated for future planning (Rainwear 1997).
In Section 2, the methodology of the proposed method is presented. The characteristics of regression models are explained in section 3. The numerical examples and discussions demonstrate the application of the proposed method is presented in section 4. Conclusions and results are summarized in section 5.

2. PROPOSED METHOD

The common and important task that faces industrial area is to determine the existence, nature and the relationship between variables. The regression analysis is widely used for this purpose. The proposed method can be described as shown in Fig.1 and structured as follows:

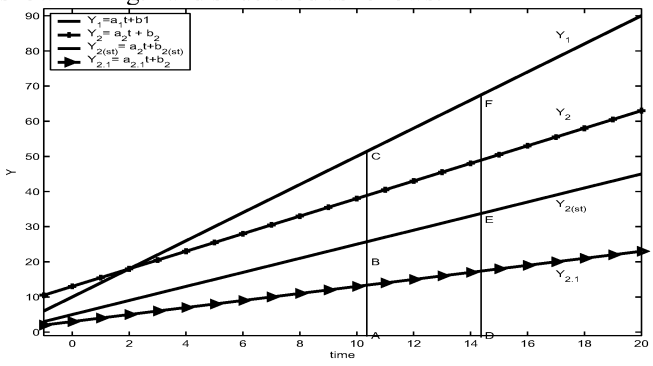


Fig.1: The level of change between regression lines

Where:

- Y – The regression line (Y = a.t + b).
- a – The slope of the line (namely, the change in Y which accompanies an increase of one unit of time (t).
- t – Time, it is equal 1 for the first year of the analyzed period and so up to n-years.
- b – The Y-intercept (the value of Y for t = 0).
- $b_{2(st)}$ – Intercept of the regression line $Y_{2(st)}$, calculated from the given values of the parameters a_1, a_2 and b_1 of the regression lines Y_1 and Y_2 (where: $Y_2 = Y_{2.1}$ to $Y_{2.n}$)
- $Y_{2(st)}$ – The standardized line ($Y_{2(st)} = a_2.t + b_{2(st)}$) with a constant level of change to line Y_1.

In order to express the nature of the said method, the level of change between the regression lines in Fig, 1 has been investigated, here, Y_2 (where: $Y_2 = Y_{2.1}$ to $Y_{2.n}$) to the regression line Y_1 and the following cases are discussed:

2.1 The level of change between two regression lines Y_1 and Y_2 is constant, if the relationship between their regression lines is constant. The following relation can be derived from Fig.1 .

$$\frac{AB}{AC} = \frac{DE}{DF} = const \qquad (1)$$

The relationship between regression lines Y_1 and Y_2 might be presented as in (2).

$$\frac{Y_{2(t=1)}}{Y_{1(t=1)}} = \frac{Y_{2(t=2)}}{Y_{1(t=2)}} \cdots = \frac{Y_{2(t=n)}}{Y_{1(t=n)}} = const. \qquad (2)$$

Wherevever , the above relation is realized, obtain :

$$\frac{a_1}{a_2} = \frac{b_1}{b_2} \qquad (3)$$

For Y_2 (t=1,2,…, n) is realizing $Y_{2\,(st),}$ Obtain:

$$b_2\,(st) = b_2 \qquad (4)$$

This means, the level of change between the regression lines Y_2 (where: $Y_2 = Y_{2.1}$ till $Y_{2.n}$) and Y_1 is constant. i.e Y_2 is equal to the standardized $Y_{2(st)}$, and the intercept b_2 becomes as given in (4). However if (3) does not obtained, then $b_{2(st)}$ can be calculated from (3) and (4):

$$b_{2(st)} = \frac{a_2}{a_1} \cdot b_1 \qquad (5)$$

2.2 If the regression line Y_1 has a steeper slope and lower intercept than the regression line Y_2. Here, b_2 is more than b_1 while a_2 is less than a_1 (as Y_2 in Fig. 1). In this case, equations (2) and (3) are not attained. Thus, the level of change between the regression lines Y_1 and Y_2 will be constant as in (2), if the intercept of the regression line Y_2 is equal to (5). For this case obtain:

$$b_{2\,(st)} < b_2 \qquad (6)$$

This means, the level of change of the regression line Y_2 to the regression line Y_1 (Fig. 1) is higher than the standardized line $Y_{2(st)}$ relizeng (2).

2.3 This case is exactly opposite to the previous one ($Y_{2.1}$ in Fig. 1). If the regression line Y_2 has a steeper slope and lower intercept than the regression line Y_1. Here, a_1 is less than a_2 and b_1 is more than b_2. Obtain:

$$b_{2\,(st)} > b_2 \qquad (7)$$

This means, the level of change of the regression line Y_2 to the regression line Y_1 (Fig. 1) is lower than the standardized line $Y_{2(st)}$ relizeng (2).

2.4 If the regression line Y_1 has a steeper slope and intercept than the regression line Y_2 or vice versa. the level of change of the regression line Y_2 to the regression line Y_1 will depend on Y_1 and Y_2 parameters (i.e: a_1, a_2, b_1 and b_2). Then, $b_2(st)$ could has a value more, less or equal b_2.

$$(b_2(st) > b_2 \quad or \quad b_2(st) < b_2 \quad or \quad b_2(st) = b_2) \qquad (8)$$

In tihis cases, the level of change of the Regression line Y_2 will equal the level change of the Regression line Y_1 if the intercept of Y_2 will equal to (5).

3. REGRESSION MODELS

The least squares method is used to provide the best possible fit (Douglas 1994). So, we shall concentrate mainly on liner regression model.

$$Y = a.t + b \qquad (9)$$

However, there are situation in which data depart more or less widely from linearity, this indicates that a nonlinear equation in a logarithmic form will provide a good fit more than the straight line fit.

$$Y = a.\ln(t) + b \qquad (10)$$

In case of electrical energy data analysis. It is worth mentioned that, the adopted two regression models, linier and logarithmic models are fit of goodness; best describes the overall pattern of the data and simply can achieve the results than the other regression models (Hu et al. 1998). To identify the goodness of fitting regression model to the actual data, the coefficient of determination R^2 is provided. The R^2 value is varying between 0 and 1. The more R^2 is close to 1, the more strong relationship is in the fitting of the data (Freund et al. 1992).

.

4. NUMERICAL EXAMPLES AND DISCUSSIONS

To establish the effectiveness of the proposed method, a study was conducted to the data available from the annul report of the National Electric Power Co. NEPCO (Annual Report, 1996 - 2003 and draft of 2004.

4.1 Generation losses.

The balance of generated energy is shown in table No.1.

Table No.1. Generated energy and generation losses [GWh].

Year	Generated Energy	Sent Out Energy	Losses	Losses%
Sn.	A1	B1	C1	D1
1991	3485	3255	230	6.60
1992	4168	3895	273	6.55
1993	4551	4256	295	6.48
1994	4837	4543	294	6.08
1995	5365	5053	312	5.82
1996	5801	5470	331	5.71
1997	6042	5716	326	5.40
1998	6520	6120	400	6.13
1999	6852	6392	460	6.71
2000	7125	6639	486	6.82
2001	7349	6856	493	6.71
2002	7828	7308	520	6.64
2003	7679	7175	504	6.56
2004	8449	7922	527	6.24

The effect of applying the regression models, linear, and nonlinear models, and their determinations (R^2) using least squares method are shown in table 1.1.

Table 1.1 Regression models for the data from table 1.

1.	Generation	Regression Model			
		linear	R^2	Non-linear	R^2
A1	Generated Energy	Y=357.5.t+3465.3	0.985	Y = 1877.9Ln (t) + 2767.4	0.930
B1	Sent out energy	Y=332.99.t+3259.7	0.983	Y = 1754.5Ln (t) + 2600.1	0.934
C1	Losses = A1-B1	Y=24.508.t+205.55	0.944	y = 123.42Ln (t) + 167.27	0.820
D1	Losses [%]	Y=0.022.t+6.1529	-0.045	Y = 0.015Ln (t) + 6.291	0.001

The relationship between the generated energy and generation losses are shown in Fig. 2.

The level of change between the generated energy and generation losses can be evaluated as a relationship of change between two regression functions, as shown in table 1.1 and presented in Fig.2:

A1= f(C1) - in liner form.

Where:

$A_1 = Y_{Gen.En.} = 357.5.t+3465.3 = Y_1$

$C_1 = Y_{Gen\ Los.} = 24.508.t+205.55 = Y_2$

Evaluating the above functions by investigating their parameters, this case is similar to the 2.4. (Where: $a_1 > a_2$ and $b_1 > b_2$).

Substituting the above data into (5):

$$b2(st) = \frac{24.508}{357.5} \cdot 3465.3 = 237.559 > b2 = 205.55$$

And , for A1= f(C1) - in non-liner form, obtain:

$$b2(st) = \frac{123.421}{375.5} \cdot 2767 = 909.60 > b2 = 167.27$$

The obtained relation is as in case 2.3, where: $b_{2\,(st)} > b_2$.

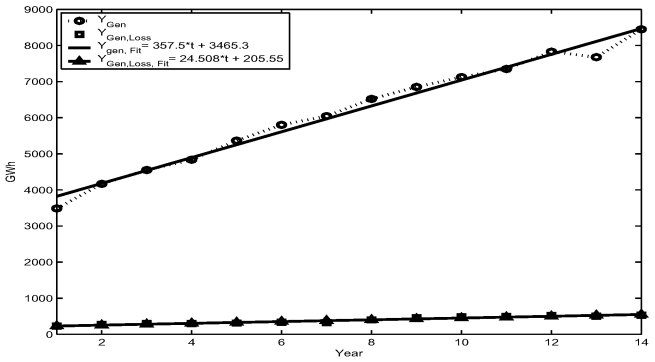


Fig. 2. Generated energy and generation losses.

This means, the level of change of the energy losses Y_2 is lower than the level of change of the generated energy Y_1. The energy losses level has a trend to increase, but less than the increment level of the generated energy.
Due to the low value of the determinations R^2 of the non-linear regression, our presentation on the scatter diagram will be limited to the straight line regression only.

4.2 Transmission losses

The transmitted energy and the transmission losses data are shown in table No.2. The level of change between the transmitted energy and the transmission losses can be evaluated as a relationship between the regression lines A2= f(C2). Substituting the above data in (4) (5), yields:

A2= f(C2) – in linear form:

$$b2(st) = \frac{21.764}{392.1} \cdot 2681.2 = 148.823 > b2 = 5.6538$$

and A2= f(C2) – in nonlinear form:

$b_{2(st)} = 113.232 > b_2 = 8.669$

This is similar to the case in 2.3, where: $b_{2\,(st)} > b_2$.

Table No.2 transmitted energy and transmission losses [GWh].

year	Sent out Energy 132 kV)1*	Bulk sales	Losses = A2-B2	Losses[%]
Sn.	A2	B2	C2	D2
1991	3082	3016	66	2.14
1992	3634	3565	69	1.90
1993	3928	3857	71	1.81
1994	4141	4074	67	1.62
1995	4612	4503	109	2.36
1996	5037	4906	131	2.60
1997	5293	5163	130	2.46
1998	5705	5562	143	2.51
1999	6270	6081	189	3.01
2000	6535	6321	214	3.27
2001	6897	6642	255	3.70
2002	7436	7129	307	4.13
2003	7967	7664	303	3.80
2004	N/A)2*	N/A	N/A	N/A

)1* Does not include industrial companies network.
)2* Data from the source was not ready yet, during the preparation of this paper.

Table No. 2.1. Regression models for the data from table No. 2.

2	Transmission Losses*	*Regression Model*			
		linear	**R^2**	**Non-linear**	**R^2**
A2	Sent out energy 132 kV	Y = 392.1.t + 2681.2	0.995	Y = 1862.4Ln(t) + 2195	0.867
B2	Bulk Sales	Y = 370.34.t + 2675.6	0.995	Y = 1766.3Ln(t) + 2203.7	0.875
C2	Losses =A2-B2	Y = 21.764.t + 5.6538	0.925	Y = 96.075Ln(t) - 8.6695	0.697
D2	Losses [%]	Y = 0.1914.t + 1.3765	0.854	Y = 0.8227Ln(t) + 1.289	0.609

The relation between the sent out energy 132 kV and the transmission losses is shown in Fig 3.

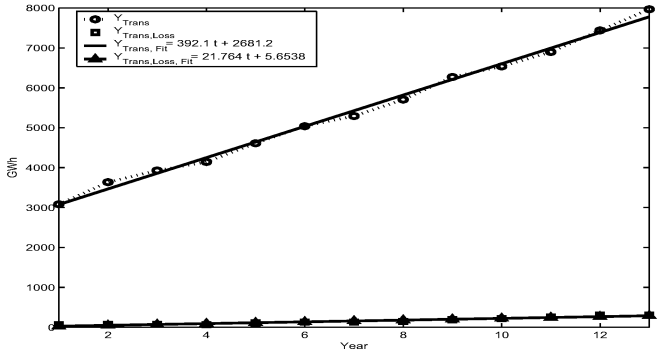


Fig. 3. Transmitted energy and transmission losses.

This means, the level of change of the transmission losses is lower than the level of change of the transmitted energy. The transmission losses level has a trend to increase, but less than the increment level of the transmitted energy.
This indicates that, the being state of the transmission system, including transmission lines, power transformers and system equipment etc..., are able to transmit more energy not effecting the transmission system nominal capacity

4.3 Distribution losses

The data concerning the sent out energy and the distribution losses and their regression models are shown in tables Nos. 3 and 3.1 respectively.

Table No. 3. Sent out energy and distribution losses [GWh]

year	Sent out Energy	Sold Energy	Losses = A3-B3	Losses[%]
Sn.	A3	B3	C3	D3
1991	2424	2163	261	10.77
1992	2776	2474	302	10.88
1993	3102	2757	345	11.12
1994	3448	3083	365	10.59
1995	3768	3373	395	10.48
1996	4153	3710	443	10.67
1997	4363	3889	474	10.86
1998	4813	4229	584	12.13
1999	5336	4758	578	10.83
2000	5646	5038	608	10.77
2001	6026	5366	660	10.95
2002	6405	5701	704	10.99
2003	6923	6113	810	11.70
2004	N/A)1*	N/A	N/A	N/A

)1* Data from the source was not ready yet during the preparation of this paper.

Table No. 3.1 Regression models for the data from table No.3

3	Distribution Losses	Regression Model			
		linear	R2	Non-linear	R2
A3	Sent out Energy	Y = 369.37.t + 1967	0.996	Y = 1748.6Ln(t) + 1519.2	0.862
B3	Sold Energy	Y = 326.51.t + 1764.7	0.996	Y = 1547.5Ln(t) + 1365.8	0.864
C3	Losses = A3-B3	Y = 42.857.t + 202.23	0.978	Y = 201.08Ln(t + 153.4	0.832
D3	Losses [%]	Y = 0.045.t + 10.665	0.147	Y = 0.1904Ln(t) + 10.65	- 0.107

Applying the above regression data of the sent out energy and distribution losses to (5), obtain:
A3=*f*(C3) - in linear form:

$$b_{2.(st)} = \frac{42.857}{369.37} \cdot 1967 = 228.225 > b_2 = 202.23$$

and A3= *f*(C3) - in nonlinear form:
$b_2(st)=174.70 > b_2=153.4$
The relation between A3=*f*(C3) is shown in Fig. 4.

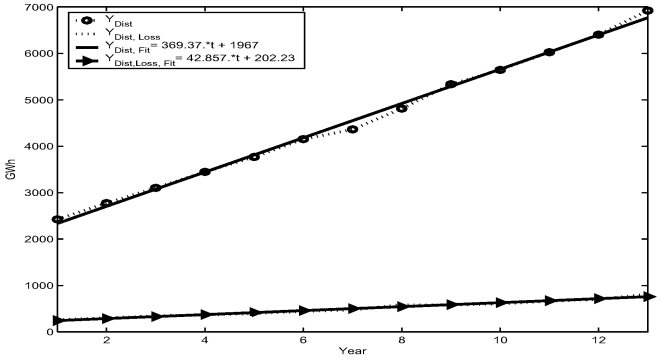


Fig. 4. Sent out energy and distribution losses [GWh]

This case is similar to the case as in 2.3, where: b_2 (st) > b_2. The level of change of the distribution losses is lower than the level of change of the sent out energy. This means that, the distribution system still able to distribute electric power with reasonable level of losses not reaching the level of change of the distributed energy. This parameter indicates the capability of distribution equipment (i.e. distribution network, transformers itc.).

4.4 Commercial (Black) losses

Commercial losses can not be considered as technical losses. Defined, as a portion of energy not tariffs precisely or entirely not tariff which, can be classified as follows (Bassaikhan et al.2003):

- Losses cussed by operators (human losses). That is occurring during the reading of energy meter or /and translocation of data to the computer system.
- Energy power meters / transducer errors or failure.
- Illegal consuming of the energy.

Total commercial losses can be determined as a difference between the summation of energy losses arising from generation losses (C1), transmission losses (C2) and distribution losses (C3) to the energy losses arising from the balance of the total system losses as shown in table No. 4.

Table No. 4 Summation and balanced energy losses [GWh].

Losses Sum		Total system losses Balance					
year	(Gen.+ Tran+ Dist) losses	Generated Energy	Sold Energy	Los.= B4-C4	Losses [%]	Commercial losses =A4-D4	|commercial Losses| = | A4-D4 |
Sn.	A4	B4	C4	D4	E4	F4	G4
1991	557	3485	2918	567	16.27	-10	10
1992	644	4168	3509	659	15.81	-15	15
1993	711	4451	3832	619	13.91	92	92
1994	726	4837	4107	730	15.09	-4	4
1995	816	5365	4525	840	15.66	-24	24
1996	905	5801	4876	925	15.95	-20	20
1997	930	6042	5079	963	15.94	-33	33
1998	1127	6520	5425	1095	16.79	32	32
1999	1227	6897	5654	1243	18.02	-16	16
2000	1308	7170	5872	1298	18.1	10	10
2001	1408	7616	6217	1399	18.37	9	9
2002	1531	8150	6629	1521	18.66	10	10
2003	1617	8651	7021	1630	18.84	-13	13

Commercial losses can be explained as a violation, which might have positive or negative values. Thus, the absolute value of the commercial losses is considered for the evaluation. These values and their regression models are stated in the below table No. 4.1.

Table No.4.1 Regression models for the data from table No.4.

Sn	Regression model	Regression Model			
		linear	R2	Non-linear	R2
A4	Losses Sum (Gen.+Tran+Dist)	Y = 89.962.t + 409.2	0.9812	Y = 416.21Ln.(t) + 316.97	0.811
B4	Generated energy	Y = 408.51.t + 3229.2	0.9964	Y = 1974.5Ln.(t) + 2663.3	0.899
C 4	Sold Energy	Y = 317.91.t + 2825.7	0.9932	Y = 1555.9Ln.(t) + 2352	0.919
D4	Losses Sum.(balance)	Y = 90.593.t + 403.46	0.9824	Y = 418.66Ln.(t) + 311.33	0.810
E4	Losses Sum(balance) %	Y = 0.3415.t + 14.334	0.7436	Y = 1.3846Ln.(t) + 14.322	0.472
F4	Comm. Losses	Y = -0.6319.t + 5.8077	0.0058	Y = -2.457Ln.(t) + 5.647	0.004
G4	| Comm. Losses |	Y = -0.9956.t + 28.538	0.0351	Y = -1.2108Ln.(t) + 23.25	0.002

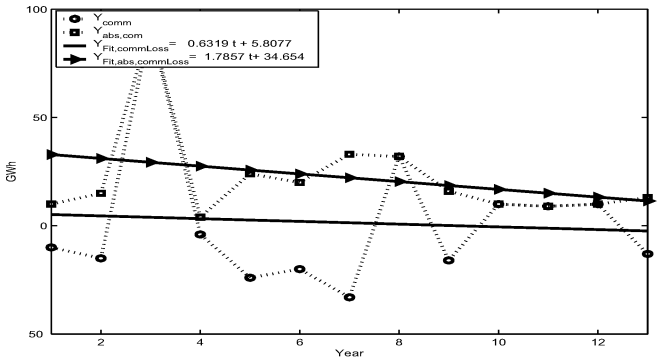


Fig. 5 Commercial losses (actual and absolute values).

The level of change between the total system losses and the commercial losses can be represented as:
E4=*f*(G4) - in linear form:

$$b_{2cal.} = \frac{-0.9956}{90.593} \cdot 403.46 = -4.43395 \langle b_2 = 28.538$$

and E4=*f*(G4) - in nonlinear form:
b_2(st)= -0.06724 < b_2=23.25

This means, the level of change of the commercial losses is higer than the level of change of the total system losses.
Due to the small amount of determination R^2 of the commercial losses, the achieved result could be unconfident. To achieve a satisfactory approximation of the commercial losses, the regression line should be presented by polynomial function to the order not less four.
Therefore, the trend of the commercial losses shall be evaluated from the relation between commercial losses and the total system losses (G4/D4 tab. 4 and in Fig. 7), which shows that, the taken action in this regard for the last 5 years brings the commercial losses from 2 to 3 percent to the value less than 1% .

4.5 Total System Losses

The level of change between the generated energy and the total system loses can be evaluated as bellow and shown in Fig. 6.
B4=*f*(E4); in linear form:

$$b_{2.(st)} = \frac{90.593}{408.51} \cdot 3229.2 = 716.1218 > b_2 = 403.46$$

And, B4=*f*(E4) as a nonlinear form:
b_2(st)=564.70 > b_2=311.33

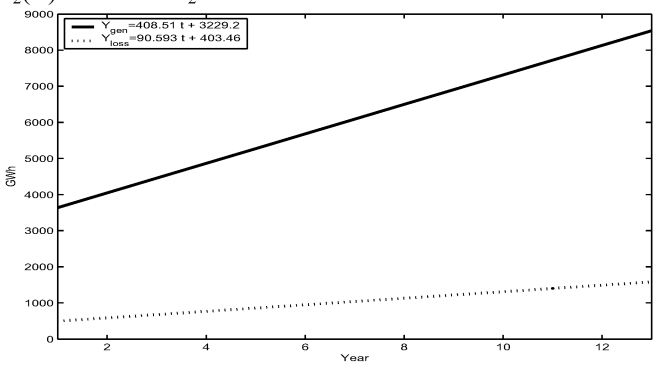


Fig. 6 Generated energy and total system losses

This means, the level of change of the total system losses is lower than the level of change of the generated energy, and any increment of generation, the system losses will increase also, but with a level of change less than the generated energy.

The total percentage system losses can be summarized and presented as in Fig.7.

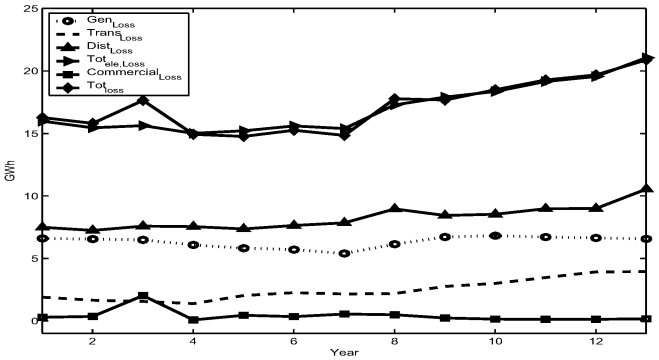


Fig.7 System Losses

The regression lines of the demonstrated data can be used for system planning and forecasting. For example, the generation energy value at the years 2005, 2010, 2015 and 2020 can be predicted from the regression line of the generated energy in table No.1.1. Using the linear form obtains:
Y=357.5.t+3465.3
The predicted values of the generation energy will be 8827.8 [GWh], 10615.3 [GWh] 12402.8 [GWh] and 14190 [GWh] receptively

5. RESULTS AND CONCLUSIONS

This paper presents the application of the regression analysis for the electrical energy data. Two regression models are adopted, linear and logarithmic, using least squares method. The linear method is more goodness of fitting regression model to the energy data than the logarithmic model. Exception is the percentage energy losses, which has a weak relationship to both models due to the low value of the determination. Thus, the result might be unacceptable. In addition to the generation, transmission, and distribution energy losses, commercial black energy losses is defined, computed and analyzed.
The proposed method in this paper has an advantage over traditional methods, that investigates the electrical energy data behavior for a long period based on historical data. The achieved results and trends can provide a good tool for the evaluation, prediction and investigate the future planning of the electrical power system.
Nevertheless, the generation, transmission, distribution and the total system losses' have a level of change lower than it should be, the losses overall the system still have a level higher than the intentional losses level. Therefore, the losses should be reduced overall the system sections. The high amount of the distributions losses to the total system losses is due to the bulk size of this network with comparison to the total system network, thus more care and investigation shall be directed to this network

REFERENCES

Agha Ayman, "Probabilistic approach for reactive power compensation", Dirasat, Engineering Sciences Magazine, Vol. 26, No.2, Oct. 1999.

Bassaikhan D, Sodonomdori D, Enklargal Khm, Narantuya M,.Dologorsuren L "Commercial losses of UB electrical distribution network shareholding company (CO., SH) and their reducing" Proceeding of the 7th Korea – Russia International Symposium, Korus2003,pp 38-44.

Chang Yung-Chung, Wei-Tzen, Chun-Chang Liu, "A new method for calculating loss coefficients", IEEE Trans. on Power Systems, Vol.PAS-9, No.3, Aug. 1994, pp 1665-1671.

Freund John E, Gary A. Simon "Modern elementary statistics", Eight Edition. Prentice– Hall international, INC. New Jersey 1992.

Glamocanin. V, "Optimal loss reduction of distribution networks", IEEE Trans. on Power Systems, Vol.5, No.3, Aug. 1990, pp. 774-782.

Karakatsanis T.S,.Hatziargyriou N.D, "Probabilistic constrained load flow on sensitivity analysis", IEEE Trans. on Power Systems, Vol. PAS-9, No.4, Nov. 1994, pp 1853-1860.

Montgomery Douglas C., Runner George C., "Applied statistic and probability for engineers", John Wiley & Sons, Inc. 1994 Toronto, Canada.

National Electric Power Company (NEPCO), "Annual Report", 1996, 1997 and 2003, Amman -Jordan.

Neagle N.M., Samson D.R., "Loss reduction from capacitors installed on primary feeder", AIEE Trans. on PAS, part 3, Vol.75, 1956, pp 950-959.

Nelson E.Chang, "Determination of primary feeder losses", IEEE Trans. on PAS, Vol. PAS-87, No.12, Dec. 1968. pp 1991-1994.

Rainwear D.K., Karady G.G., Farmer R.G. "Economic impact analysis of load forecasting", IEEE Trans. on Power System Vol. 12, No 3, Aug. 1997, pp 1388 1392.

Wu C.J, Hu C.H., Yen S.-S, Yin C.-C., Chin C.-C, Lee Y.-M, "Application of registration models to predict harmonic voltage and current growth tread from measurement data at secondary substations", IEEE Trans. on power Delivery, Vol. 13, No 3. July 1998, pp 793-799.

AUTHORS BIOGRAPHY

Nabeel I. Tawalbeh was born in Jordan in 1957. He received his B.Sc. and Ph.D. degrees in electrical engineering from the University of Damascus- Syria and Warsaw University of Technology, Poland in 1985, 1997 respectively. He works as instructor engineer at Polytechnic Alhuson. 1992 to 1997 delegated to Poland for Ph.D. After that from 1997 to 1999 he worked as lecturer at Yarmouk University Jordan. Currently he is serves an assistant Prof. at the Department of the Electrical Engineering at Al-Isra Private University-Jordan. His interests include power systems analysis, Load estimation and forecasting, and statistical approaches for electrical systems. nabiltawalbeh@yahoo.com

Ayman Agha was born in Irbid, Jordan in 1961. He received M.Sc. and Ph.D. degrees in electrical engineering from the University of Mining and Metallurgy in Cracow, Poland in 1989 and 1997 respectively. He joined Jordan Phosphate Mines Co. (JPMC) in 1989 as maintenance, Planner, Preventive maintenance engineer and at the project's department till 2000 year. From 2000 to 2004 he worked as a Studies and Design Dept. Manger in alfanar Construction Co in Riyadh - KSA. In 2004 rejoined his previous work in (JPMC). His interests include power systems, nonlinear systems, optimization theory and reactive power compensation. aay@o2.pl

LOGISTICS SIMULATION

REVERSE LOGISTICS OF END-OF-LIFE VEHICLES: MODELLING AND SIMULATION OF COLLECTION AND DISASSEMBLY

Gabriella Maria Acaccia
Rinaldo Carlo Michelini
Nicola Qualich
PMAR*lab*- Laboratorio di Progettazione e Misure per l'Automazione e la Robotica
DIMEC- Dipartimento di Meccanica e Costruzione delle Macchine - Università di Genova
Via all'opera Pia 15°- 16145 Genova – Italy
{acaccia | michelini | qualich}@dimec.unige.it

INTRODUCTION.

The critical nature of growth sustainability pushes to reverse logistics processes, requiring new knowledge references, to be developed by specially implemented information processing tools, notably, occurrence-driven simulators. The emerging businesses paradigms are recalled, in that generic frames, and exemplified through the vehicles end-of-life problems, with focus on the EU regulation and on the case-situations of different national contexts. The description summarises the backward material flow and joined information processing to characterise, both, physical and logical entities, duly accounted by the specially implemented simulation environment. For explanatory purposes, the presentation ends, discussing example issues, obtained by simulation.

KEY WORDS: End-of-Life Vehicles. Withdrawal and Disassembly.

STATE OF THE ART.

Environment-friendly behaviour and sustainability.

The affluent society, brought forth by the industrial revolution through the economy of scale, cannot last for long, being based on the ceaseless replacement of goods, manufactured by depleting earth resources, transformed into pollution and wastes. Eco-consistency requires new business paradigms. The measures for sustainable growth are, possibly, not fully defined, and to address closed-loop supply chains is, only, an attempt to re-establish more conservative trends to the thrifty society. Indeed, the earth is an almost isolated system, with (limited) energy taken from sunlight and released into space, and this is an energy, not entropy, balance **[1-4]**. In time, the eco-decay will limit growth, unless up-grading ecological issues are found, non dependent on entropy, e.g., by bio-mimesis or restoration emulating the living beings.
Recycling becomes challenging demand. Beside technical aspects, spur is: legal acts (suppliers responsibility, landfills regulation, etc.) *vs.* economic profits (recovering, remanufacturing, etc.). The new business, indeed, is widely stimulated by law-enforced recovering quotas and taxation policies in commodities procurement, or by waste disposal restrictions and correspondingly high fees. Thereafter, the envision of technology-driven solutions shall move with factual checks of their financial appropriateness. The emerging industrial fields will track reciprocal ways **[5-8]**:

- to modify the forward value chain, expanding the intangibles delivery for more efficiently satisfying clients' needs, while watching the eco-impacts; profitability is achieved by selling *products-services*;
- to resort systematically to the backward value chain, for effective utilization of the pawned natural capital; profitability is deferred to *reverse logistics* activities, rewarded by taxing actual net burdens.

Reverse Logistics.

The *reverse logistics* is the process of designing, planning and controlling the return and re-use of worn-out products, in order to conserve resources and protect the environment **[9-12]**. Profitability through *reverse logistics* requires boosting facts, such as the following:

- fully operating closed loop supply chains, grounded on standards for parts/materials recovering;
- tendency to remanufacture, based on original design for reuse, instead of native disposal destination;
- effective collection and disassembly organisations, with effective facilities for worn-out goods;
- improved sorting, retrieval and reintegration processes, to extract assessed quality resources;
- account of material and energy provisions on the supply chain, to repay the net caused decay;
- recognition of every emerging impacts, to pay for the remediation and restoring incumbents;
- exhaustive information frames, for product data lifecycle certified monitoring and vaulting;
- objective tax collection for all natural capital withdrawals, to establish fair-trade competition.

These are a mix of legal and technical prescriptions. Anyway, dropping aside longer term scenarios, the economical assessment of reverse logistics activities ought to be current challenge for the coming years, in keeping with governmental enacted rules and actual technical opportunities. To that goals, it is especially

relevant to develop decision aids for backward supply chains, and to compare collecting, dispatching, re-manufacturing and disposal processes, figuring out drawbacks/benefits each time reached. The study shall progress by means of modelling and simulation, with resort to IT aids, developing accurate virtual prototypes of the reverse logistics facilities, so that alternatives are tested in running conditions and prospected technical options are ranked as for cost and performance.

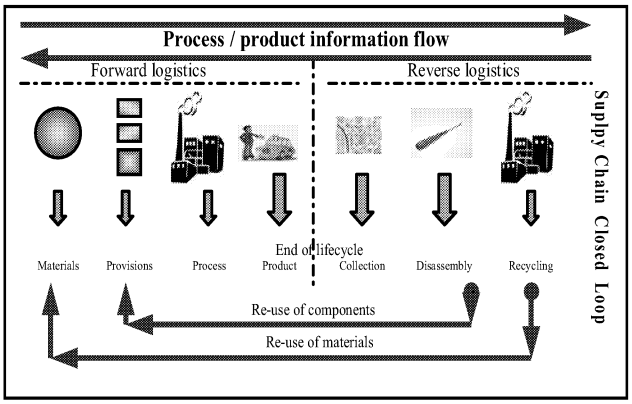


Fig. 1. Closed loop of supply chain.

On these premises, the impact of reverse logistics is differently perceived in the today world, with stress on weighing current *vs*. future, and local *vs*. global needs, according to rather opaque procedures. Economical return, basically, is standard reference to support remuneration to the whole flow: - exhausted goods collection and forwarding to suitable sites; - selective dismantling with parts recovering and residuals shredding; - second hand materials sorting and enhancement, before regulated dumping. The single steps have consignors and assignors, but the outcomes do not have explicit purchasers, lacking any link to actual needs. Thereafter, the separation of backward, from forward chain will authorise deviating issues, to keep benefits within a subset actors, while leaving drawbacks to third (and future) parties. The cost/profit ratio is, accordingly, to be assessed balancing the advantage of the consumers side (manufacturers and users), *against* the protection of the all people not involved by the specific supply chain. In that sense, regulation for eco-conservativeness is peremptory imperative that governmental organisations need undertake, enacting sets of rules for charging the consumers side for the whole environmental falls-off linked to each supply chain, dismissal and recycling incumbents included. These is leading the EU Council to expand the suppliers responsibility to the points of withdrawal of basic mass products (white and brown household-appliances, automotive vehicles, etc.), as shared commitment, established at the point of sale, when sellers earn for their work, and clients satisfy their whims.

The European ELV regulations.

The end-of-life vehicles regulation provides explanatory hints to this new trend. The 2000/53/EC Directive, enacted the 18.09.2000, and requiring acknowledgement by the Member States before April 2002, defines the rules to be followed by national organisations for vehicles dismissal, and by automotive manufacturers for selling cars respectful of the legal requests. The Commission further modified and integrated the rules by special specifications, e.g. the 2003/138/EC, to establish standard codes for parts and materials. It needs to be mentioned that the European Commission is concerned by *final issues*, not about *how* these achievements are obtained according to the different Member States laws. Actually, this creates drawbacks to manufacturers, which ought to comply with possibly uneven interpretations; then, a Guidance Document is added, to collect proper harmonisation suggestions. The overall documentation, fast growing and, by now, with open doubts, can be found in **[13-16]**.

The reading of the documents leads to few comments, such as the following:

- regulation addresses environment protection, through explicit producers responsibility charges;
- design for disassembly, for recovering, for recycling, etc. practices become standard requirement;
- extensive resort to re-manufacturing, to re-use of parts, materials, etc. is promoted and stimulated;
- appropriate aids (modularity, identifying codes, etc.) are suggested to make easier dismantling, etc.;
- prescribed lifecycle monitoring and reporting disclose on-duty conformance-to-specification checks;
- not-justified high-impact, non-consistent behaviours are heavily taxed or totally forbidden.

The sample notices show that new business paradigms are fostered, based on different design patterns, with high concern on point-of-service performance and full commitment for withdrawal. The sharper focus on growth sustainability is basic driver, and, obviously, it leads to reverse logistics entrepreneurship, with larger role for competitiveness played by the incumbents of the backward cycle. We are facing one of the many turning points of the industrial evolution, to aim at explicit costing of resource depletion and environment decay. These extra-charges are, traditionally, felt unproductive investment and such they are, when the consumers side (manufacturers and users), only, is considered, as if the natural capital could ceaselessly be exploited, irrespectively of the down-grading suffered by third parties (and future population). Of course, sub-sets of companies and communities could try to postpone or remove the incumbents; this will, more or less, result in choices such as to delay investments in flow-shop factories, when scale economy was winning paradigm, or the ones in flexible automation, when scope economy turned to be competitive push.

Today, the governmental authorities should analyse the whole supply chain, which distinguishes one product from another, as for the forward and the backward track, suitably defining the pertinent taxing rule. The final costing will be ascribed to the consumers side, with tax collection duty, mainly, put in charge to the goods suppliers, at the point-of-sale. The automotive field is paradigmatically noteworthy case. It represents

a widely spread domain of *registered* goods, thus: falls-off affect a large amounts of people; and items end-of-life have individual acknowledgement and recording.
The mentioned EU approach, by the already enacted directives and advices, gives detailed prescribing issues for the end-of-life vehicles operations, such as:

- member states shall establish collecting systems for exhausted vehicles and parts, at authorised sites, where preliminary treatments will grant safety and security fitting out, by removal of noxious stuffs and harmful parts;
- withdrawal needs to be assured without charge on the final owners (prescription to be fully enabled from 01.01.2007), but included in the product-service delivery, as inherent attribute;
- dealers (and manufacturers) shall bear complete responsibility for the delivery lifecycle, dismissal included, answering for the environment impact and resource consumption according to established schemes;
- users co-responsibility could be invoked, for special non-conservative behaviours, when critical pieces are damaged, removed or modified, altering the original setting of the supply;
- the dismantling and destruction incumbents ought to be certified, with full assessment of recovery parts, re-cycled materials, thermo-recovery and residuals dumping, to be notified to the European Commission.

The prescribed issues are quite severe, say:

- from 01.01.2006: 85% by weigh of the vehicle ought to be recovered or recycled, 10% can be dumped to landfills (after suitable neutralisation) and 5% can be used as auxiliary fuel;
- from 01.01.2015: figures are modified, allowing 10% for fuel use, but only 5% to landfills.

The EU prescriptions have differently been transferred to the national regulations, and some country is still considering transient alternatives for legal (data registration and vaulting, etc.), and for technical (withdrawal, dismantling, recycling, dumping, etc.) incumbents. As a matter of facts, the ecological concern will be translated into specified goals; the local competitiveness will outcome from the overall effectiveness enabled by individual governmental organisations, from the capability of the regional service-enterprises (cars repair shops or scrapers, spare parts dealers, etc.) to efficiently respond, and from automotive industries motivated to fill the emerging business with competitive offers. The role of latter actors is fundamental, at this stage, to support reverse logistics by product-data management aids for enhanced dismantling and selective recycling, and turns to be critical, in the short time, to enable design-for-recycling paradigms (modularity and material segregation, disassembly pre-setting, etc.) and to take advantage of sustainable manufacturing (high resort to recovered or recycled provisions, etc.) and maintenance (re-integration, proactive up-keeping, etc.) options. The automotive business, indeed, presently suffers of strong world-wide competition and over-production; firms not quickly reacting to the new requests, shall most likely become marginal and will be moved off the market.
To get the whole event in perspective, few figures give the size of the automotive reverse logistics business, at the EU regional basis. In 2001, about 225 millions vehicles (some 185 millions cars) where circulating in the 15 Countries (say, a vehicle each two inhabitants). These figures show the approaching of saturation, and the substitution-based market of the ten to fifteen years old cars leads to a reverse logistics business of some 10 millions items per years. There is room for economy of scale factories, replacing today car breakers.

The ELV reverse logistics chain.

Looking after the basic prerequisites, a few facts characterise the emerging business processes of ELV reverse logistics:

- the need of a logistic network, with collecting and transportation equipment, storage points, handling and processing fixtures, and joint information flow for acquisition and recording;
- the establishment of adequate dismantling shops for safe parts recovering, and of suitable shredding facilities, to grind the left hulks to tiny pieces;
- the resort to specialised sorting plants, to separate the different metals (ferrous alloys, stainless steels, brass, aluminium alloys, etc.) from glass, plastics;
- the development of recycled materials tracks, with enhancement of design and manufacture with secondary materials provisioning practices.

All four processes lead to innovative set-ups; the second and third ones deserve particular interest, with new robot aids to be conceived and exploited for irksome and dangerous tasks, today, mainly accomplished by front-end personnel. It should be noted that the material flow, as a hole, deals with a poor, but heavy, stream; the cost of handling, carrying and dispatching could become serious drawback, with high energy consumption. Effectiveness will, probably, look after distributed processing units, characterised by suitably innovative technologies. In the overall backward cycle, however, the joined information flow bears a critical relevance. Actually, it is worthy to distinguish:

- the reference knowledge (IDIS, etc.), to make efficient the dismantling, recycling, etc. duties, according to the product data compulsorily provided by the manufacturers;
- the operation flow, with data acquisition, handling, processing and vaulting, to issue the requested records and certificates and give account the process compliance;
- the framework assessment, to evaluate competing arrangements and to develop high yield, availability and safety reverse logistics facilities.

The present study specifically addresses the latter opportunity, and considers the *simulation aids* as basic means, to be used, at the design and development stage, to establish the conditioning knowledge as decision support to chose the reverse logistics facilities,

and, at the on-duty operation stage, to reset the process plans to optimise productivity and efficiency, for the actual running agendas.

PMAR*RLelv*: ENVIRONMENT TO SIMULATE ELV REVERSE LOGISTICS.

Moving in such a context, the simulation aids are developed, exploiting occurrence-driven architecture, to provide real-time, dynamical description of backward processes, directly duplicating the effects of every forcing (and disturbing) inputs, as they appear in the reality. At the same time, alternatives of physical resources and/or logical plans can be implemented, making un-expensive virtual tests, without actually building pilot plants. The basic options offered by this kind of simulation aids are well known, and we do not enter into details.

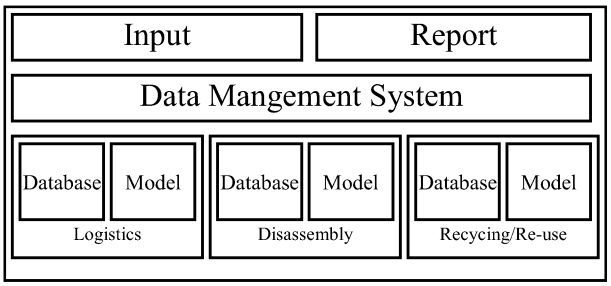


Fig. 2. PMAR*RLelv:* an environment for simulation of ELV reverse logistics.

The PMAR***RLelv*** [Fig. 2] environment has the featuring property to be based on standard packages:

- **WITNESS** software **[17]** version *logistics*, for the collection-dispatching modules, and version *manufacture*, for the dismantling-shredding-sorting modules, with virtual reality aids, leading to self-sufficient and friendly modelling . Parametric modelling is obtained by hierarchic and modular architecture; for each block, the behavioural logic is coded by rules, concerning input/output schedule, occurrences chaining, and decision setting.
- **MS ACCESS** software by Microsoft, being well known relational database, released as MS office suite.
- **VISUALBASIC.NET** software, based on the **.NET**-platform, powerful and flexible, but reasonably friendly also for web applications, assuring properly effective management aids to create and run interfaces with other languages, procedures and databases; to establish masks and graphic displays; to define interactions between blocks ; to start or stop concurrent processes; to enable web applications.
- **CRYSTAL REPORT XL** software **[18]**, for providing standard frames to reporting.

Basically, the availability of the PMAR***RLelv*** environment aims at three scopes:

- the creation of digital mock-ups of actual physical resources involved in reverse logistics duties;
- the provision of personalised access to web-clients, interested to investigate special opportunities;
- the build-up of qualified knowledge, by systematic ranking set of architectures and schedules.

A bird eye view of the work shows the interlacing of three functional streams: transport and dispatching; dismantling and recovering and shredding and sorting, with local buffering and on-progress data handling. At the moment, the three streams bear quite different operation styles, only sharing, perhaps, lack of structured patterns: end-of-life vehicles are fetched at random times and scattered locations and dropped to off-hand wrecking shops; disassembly and reuse are often done by on-process operators, as case arises; only shredding and sorting have resort to mechanised equipment, even if human-intensive attendance is kept.

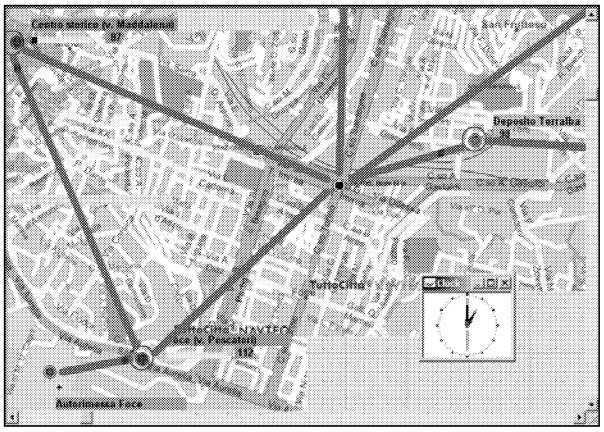


Fig. 3. Typical organisation of the input knowledge (model layout).

The simulation environment, of course, shall give account of the existing state-of-the art. The reverse logistics, thereafter, deals with checking the competitiveness of alternative transportation and delivering lay-outs and schedules, alternative break-down van fleets and buffering distributions, alternative recycling and dumping policies, alternative wrecks reduction and materials extraction facilities, and so on; the divide, between the hastily botched jobs of today, and the structured (industrialised) organisations could, on-progress, establish, once the return on investment will suggest to enable properly modified and optimised processes. As a matter of facts, the sequence shredding and sorting seems quite poor, but it happens to be *best* solution, when tight and intimate connection of parts makes impossible simpler decoupling; when, in the future, wreck reduction will be pre-set according to proper design-for-recycling protocols, the all backward flow could turn to higher performance plans. Summing up, the functional modelling for simulation is powerful means to detect strength and weakness of actual arrangements, and the comparative assessments help addressing the investigation toward the most critical aspects.

Example: collection of ELV in the town of Genova

The bylaws for car registration and end-of-life

processing differ in the different EU Countries, with emphasis on legal or technical requisites. In Italy, the ownership registration/cancellation incumbents are quite onerous, at time, exceeding the value of the transferred vehicle; this results in abandoning non-identifiable wrecks somewhere along the roads, to get rid of charges. The local boroughs need face reclamation duties, with the further constraints to satisfy the enacted rules. The PMAR***RLelv*** environment has been used to become reference tool for the local authorities of Genova, and to establish the most effective withdrawal policy for left aside end-of-life vehicles. These reclamation duties are let out to a set of eight, centrally located, wreckers, required to fetch and bring the left-out vehicles to two store-shops in the town near outskirts.

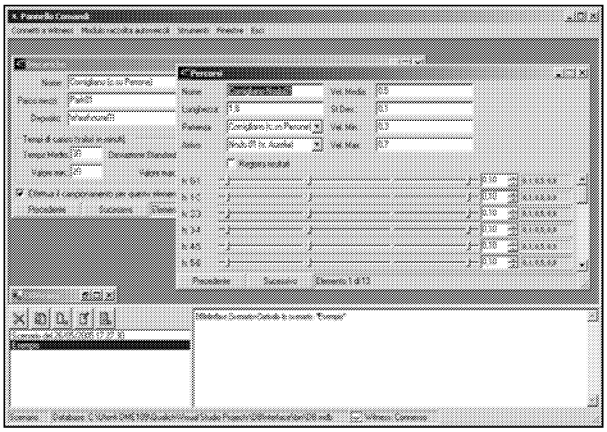


Fig. 4. Typical organisation of the input knowledge (data management system).

The problem definition [Fig. 3-4] deals with: - the map covered by the withdrawal service; - the travelled road lay-out, with weigh due to traffic periods; - the duty-service of the wreckers; - the capacity of the two store-shops. The analysis should consider several competing lay-outs, changing the wreckers allocation, the duty-cycles, the job priorities, etc.; then typical performance charts are generated, providing synthetic indices [Fig. 5-6] to show the wreckers utilisation ratio, the main characteristics of the simulated task, with specification of the total amount of the removed vehicles, and the averaged duration of each service.

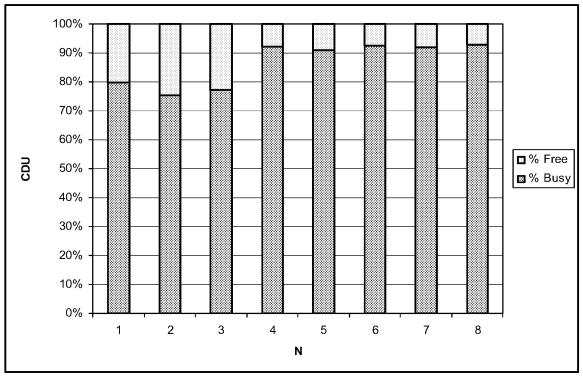


Fig. 5: Typical organisation of the output knowledge (wreckers utilisation ratio).

The withdrawal of abandoned vehicles is preliminary study in ELV reverse logistics, where transportation and handling costs exceeded dismantling and recovering ones. By simulation, rough assessments of the relative weight are quickly obtained.

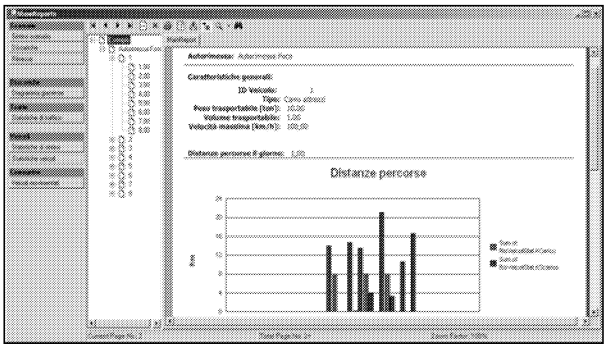


Fig. 6. Typical organisation of the output knowledge (wreckers travel statistics).

CONCLUDING COMMENTS

In the framework of the emerging area of reverse logistics, the PMAR***RLelv*** environment aims at three mentioned scopes:
to create digital mock-ups of the physical resources;
to provide personalised access to web-clients;
to systematically rank lay-outs and schedules.
The environment is being implemented based on standard software and opened to cover the many problems arising along with the all backward cycle. The recalled explanatory example covers the initial duties for the sub-set end-of-life vehicles recovery, abandoned by un-known final owners, problem specially relevant in Italy, due to local reasons. In addition, the PMAR***RLelv*** environment covers the subsequent phases of car dismantling and parts recycling, of residuals shredding and materials sorting, etc., each time modelling the selected physical process and decision logics. The study, according to the above three scopes, helps proving twofold outcomes: to stress on the relevance of the new business paradigms required by the sustainability of the growth; to organise these emerging industrial trends with full visibility of their *economic* and *ecologic* benefits and drawbacks.

Acknowlegdment.

The work is done within CIELI, *Centro Italiano di Eccellenza sulla Logistica Integrata*, established in Genova, with sponsorship and support of MIUR, Ministry of Education, Research and University, as multi-disciplinary centre, where researchers from the Business, Engineering and Law Schools work together in the domains of logistics. The PMARLab is charged to co-ordinate the domain of reverse logistics.

REFERENCES.

[1] R.C. Michelini, R.P. Razzoli: "Collaborative networked organisations for eco-consistent supply-chains", G.D. Putnik, M.M. Cunha, Eds., Virtual Enterprise Integration: Technological and Organisational Perspectives, IDEA Group, IGI Press, Hershey, PA, pp. 45-77.

[2] R.C. Michelini, G.L. Kovàcs "Information infrastructures and sustainability", L. Camarinha Matos, Ed., Emerging Solutions for Future Manufacturing Systems, Springer, pp. 347-356, 2005.

[3] R.C. Michelini, R.P. Razzoli: "Product-service ecodesign: knowledge-based infrastructures" Intl. J. Cleaner Production, Elsevier, vol. 12, n° 4, May 2004, pp. 415-428.

[4] R.C. Michelini, R.P. Razzoli: "Product-service for environmental safeguard: a metric to sustainability" Intl. J. Resources, Conservation and Recycling, vol. 42, n° 1, Aug. 2004, pp. 83-98.

[5] R.C. Michelini, R.P. Razzoli: "Affidabilità e sicurezza del manufatto industriale: progettazione integrata per lo sviluppo sostenibile", Tecniche Nuove, Milano, 2000, pp. XII-278.

[6] R.C. Michelini, F. Crenna, G.B. Rossi: "Diagnostics for monitoring maintenance and quality manufacturing", C.T. Leondes, Ed.: Computer Aided Design and Manufacturing: Techniques and Applications, vol. I, CRC Press, Boca Raton, 2001, pp. 5.01-5.66.

[7] R.C. Michelini, G.L. Kovacs: "Integrated design for sustainability: intelligence for eco-consistent products-and-services", The Estonian Business School, EBS, Review, Tallin, Winter 2002-3 issue, n° 15, Dec. 2002, pp. 81-95.

[8] R.C. Michelini, G.L. Kovacs: "Intelligent integrated design for sustainability: products-services", 5th Intl. Conf. Computer Science and Information Technologies, Ufa, Russia, Sept. 16-18, 2003, vol 1, pp. 31-38.

[9] D.F. Blumberg: "Introduction to management of reverse logistics and closed loop supply chain processes", CRC Press, Oct. 2004, pp. 248.

[10] R. Dekker, M. Fleischman, K. Inderfurth, L.N. Van Wassenhove: "Reverse logistics quantitative models for closed loop supply chains", Springer, 2004, p. 436.

[11] H. Dyckhoff, R. Lackes, J. Reese: "Supply chain management and reverse logistics", Springer, 2004, p. 426.

[12] M. Fleischmann: "Quantitative models for reverse logistics", *Springer*, Berlin, 2001.

[13] J. Wengel, P. Warnke, J. Lindbom: "Case study: automotive industry personal cars", Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, The Future of Manufacturing in Europe 2015-2020, The Challenge for Sustainability, JCR-IPTS, 2002, Sevilla.

[14] BMWGroup: "Innovation, efficiency, responsibility: sustainable value report 2003/2004", Brand: BMW, Rolls-Royce Motor Cars Limited, MINI.

[15] CEN/TC 273 "Logistics: WG4, Logistics performance measures requirements and measuring methods", Report BT N 5976, 19 November 1999.

[16] European Commission: http://europa.eu.int/comm/environment/waste/elv_index.htm.

[17] Lanner: http://www.lanner.com.

[18] http://www.businessobjects.com/products/reporting/crystalreports/start_xl.asp.

MIXED-AUTOMATION OF A SORTING POSTAL FACILITY: DESIGN AND PERFORMANCE ASSESSMENT

Gabriella Maria Acaccia
Rinaldo Carlo Michelini
Nicola Qualich
PMARlab- Laboratorio di Progettazione e Misure per l'Automazione e la Robotica
DIMEC- Dipartimento di Meccanica e Costruzione delle Macchine - Università di Genova
Via all'opera Pia 15°- 16145 Genova – Italy
{acaccia | michelini | qualich}@dimec.unige.it

ABSTRACT.

Flexible automation is powerful means, to provide effective and reliable solutions to applications, which join highly interlaced and complex flows, fluctuating work-in-progress and delivery requests, and extended mix of processed items. In these contexts, the resort to resource modularity and to mixed-mode automation (with front-end operators) needs to be investigated, to assess the return on investment, dealing with alternative lay-outs and/or schedules. Modelling and simulation become critical aids the help selecting the facilities, at the design and development stages and their control policies, at the on-duty exploitation. For explanatory purposes, the case application of the postal delivery services in taken into account, providing duty-driven example hints, with, mainly, methodological issues, with due concern on the scope economy paradigms.

KEY WORDS: Virtual Testing - Mixed-Automation Facilities - Postal Delivery Services.

INTRODUCTION.

Flexible automation is acknowledged opportunity, providing economy of *scope* effectiveness, whether fixed automation cannot be supported by economy of *scale* issues. The set-up is consistent with robotic equipment and information technology aids, assuring adaptive front ends and on-process programming. Never the less, the plant economical profit stays on with critical soundness, and careful validations are requested, before moving from concepts and design, to actual lay-outs. Indeed, today computer-aided tools and robot-based technologies are sufficiently sophisticated, to provide accuracy and versatility up to almost any need, with, of course, different costs. Flexible automation, then, shall not build on technology-driven choices, unless after checking the return on investment for actual operation conditions. This leads to a design approach **[1-3]**, heavily grounded on modelling and simulation, with beforehand setting the digital mock-ups of processing facilities and operation schedules, and extended investigation by virtual testing.

The approach supports the best practice, from concept ideation, through equipment construction and process-plan trimming, with on-duties issues transparently shown to designers. To achieve such vision, key goals are **[4-7]**:

- to describe every work-cycle, by detailed assessment of expected functions and performance;
- to specify the material flow, with characterisation of the selected front-end resources;
- to properly fix sample operation agendas, including discontinuities and unexpected occurrences;
- to establish overseeing and governing rules, fit for the programmed achievements;
- to perform simulation runs, on the (in progression chosen) equipment and process-plan.

Virtual testing aims at several scopes, including **[8-9]**:

- acknowledging the actually reached performance, to show the effectiveness of the chosen resources;
- assessing the offer reliability range, to readily enable robust set-ups, without try physical prototypes;
- evaluating lifecycle and/or by-product issues on their users' friendliness and eco-conservativeness.

The modelling and simulation codes need, accordingly, expand to reach: *completeness*, as for every scope to be tested; *consistency*, as for the incorporated relational frames; and *adequacy*, as for supplying duty checks and suggestions to innovation. These requirements lead to meta-models and to object programming aids. By the former, the detail-level is composed to create (macro) models that reflect attributes and behaviours of the all, easily propagating any change of details properties. By the latter, the knowledge is encoded by structured contexts (value, attribute, belief), assuring information inheritance through data instantiation and likelihood graduation. Happily, the available programming context offers extended means to face current cases, and software aids to help transforming models into simulation environments. WITNESS, for instance, is language specially suited for investigating occurrence-driven evolutions; it supports interactive dynamic processing, with valuable graphic restitution aids, both, for model building and behaviour displaying. Parametric models are obtained by hierarchic and modular architecture; in each block, the behavioural logic is coded by rules, concerning input/output schedule, occurrences chaining, and decision setting. The WITNESS, **[10]**, release can be expanded with optimisation and virtual reality restitution blocks.

The presentation will consider topics in investigating problems of flexible automation, where effectiveness, versatility management and return on investment

suggest careful consideration of mixed-mode operation and front-end human adaptation. The competing issues are acknowledged by developing effective modelling and simulation procedures and obtaining the backing knowledge framework. The discussion goes together with explanatory examples, specially dealing with the automation of postal services, within the range of district sorting, storing and dispatching operations. Example results are shown for methods presentation, out of sharp links with specific facility exploitation, in view of helping to design and to exploit mixed-automation plants, with effective balance of front-end operators.

Model peculiarities and paper organisation.

A sorting/storing/dispatching postal facility is a local shop for mail processing, which takes into charge sets of objects (standard letters, registered letters, folds, flats, parcels, etc.), after withdrawal from the authorised collection points (such as mail pillars/boxes, postal offices, large clients, etc.), and operates selective picking to establish separated sheaves, for proper shipment and delivery to the expected addressees. The processing distinguishes because of:

- ✓ *complexity*: manifolds with interlaced streams and highly diversified consignees and receivers;
- ✓ *condition*: unsteady flows, with unexpected surges, season periodic peaks and random fluctuations;
- ✓ *items mix*: in spite of standardisation, a wide range of objects ought to be simultaneously accepted.

The investigation needs achieves a twofold outcome, namely:

- ⇑ resources selection: work-stations, fixtures, feeders, work-cycles, front-end operators, etc.;
- ⇑ process planning: task sequencing, transport-handling programming, work-shift turnover, etc..

The work organisation policies, currently explored for manufacturing plants, are mostly unsuited for postal facilities, due to manifold complexity, very large number of processes and parts, lack of stream continuity with local surges and fluctuating averages, and request of flexible scheduling to keep the service reliability and quality figures. The process automation quickly faces computational burden. Still, economy of *scope* effectiveness shall aim at **[11-14]**:

- ⇔ dealing with the daily deliveries, properly adapting resources selection and process planning;
- ⇔ accomplishing the programmed service, with optimal exploitation of the involved facilities;
- ⇔ propagating the reliability and quality figures, also in front of localised unexpected occurrences.

The actually reached issues shall be continuously monitored and performance recorded to account for service effectiveness. The domain presents several peculiarities. Typically, postal regulations require the appropriate tracking of individually ordered batches and overseeing of the shunting and clearing duties, with vaulting of the intermediate assessments, to report current throughputs and due-dates. Performance figures include:

- # productivity, to acknowledge the actually delivered amount of processed items, on a shift term;
- # effectiveness, to assess the ratio between consignees within the due-date and actually sorted items;
- # efficiency, to describe the utilisation figure of the individual resources in the processing facility;
- # reliability, to specify the ability to keep the previous figures within appropriate boundaries.

facility analysis and lifecycle function specification:
• *description and classification of the functional areas and equipment;*
• *operation cycles assessment: constraints, parameters, performance;*
• *governing rules allotment: strategic and tactical allocation;*
model definition and choice of competing solutions:
• *processes modelling and causal blocks programming;*
• *governing logic, task assignment and decision rules coding;*
definition of the simulation environment architecture:
• *implementation of the simulation/emulation blocks and interfaces;*
• *selection of process/govern parameters and database organisation;*
• *activation of the information bases and data management aids;*
simulation-driven design operations:
• *execution of simulation runs and classification of competing solutions;*
• *preparation of adaptive improvements, linked to the candidate solution*

Fig. 1. Simulation-driven lifecycle design procedure.

The scope-conditioned design procedure follows a four steps approach, **Fig. 1**, to acknowledge the lifecycle properties. The procedure manages algorithmic and heuristic blocks; front-end operators for specific actions during mail sorting, address recognition, mismatch overruling, etc., timely adds, making possible accurate analyses of specific alternatives with mixed-automation setting, comparing each issue in terms of expected performance figures and returns on investments. The simulation tool exploits the WITNESS software **[10]**, and extensively investigates the alternation of the sorting processes, with dynamical allocation of the human resources. The governing module splits into an on-process actors (men and machines) control, and a remote supervisor for stream routing and dispatching up-grading. Example results give explanatory views of the design methodology.

Function assessment and facility analysis.

Among the applications of flexible automation, where on-process front-end operators still play relevant roles, the mail delivery deserves special concern, with important contributions for the inward and outward steps, as well as for transportation, sorting and dispatching. The evolution in the field opens many opportunities, and news business paradigms establish, since globalisation breaks earlier governmental monopolies, and several enterprises are allowed to compete, carving out niches to emerge with return on investment. This requires a quite different way to look at the performance figures, productivity, effectiveness, efficiency and reliability, distinguishing the protected postal dispatching, from the value added services, offered to satisfy very special needs of individual clients. Indeed, today mere transmission of information avails of intangible means, more suited than traditional transfer of material supports; mail delivery turns to become specialised business, when the material items themselves bear intrinsic relevance, and need be displaced and carried to the hands of the addressee. In these conditions, future scenarios should deal with progressive decay of some traditional pieces (conventional correspondence) and concern on peculiar

items (registered letters, advertising folders, etc.), in keeping with the further spur of globalisation, namely: increasing worldwide inward and outward mails, and plurality of service dealers, in parallel with the local national organisations.

MODELLING UNIVERSAL POST SERVICE.

On these premises, the automation of the mail delivery becomes challenging job, requiring careful checks at the design and operation stages: the appropriate setting of facilities and technologies cannot assure return on investment, unless efficient exploitation of resources is achieved, by on-process adapting, both, material and data flows according to bottom up requirements. In that prospect, the resulting postal service will aim at exploring *intelligent* sorters and dispatchers, with intensive resort to value added abilities, **[15-17]**. Rather than full mechanisation of all handling/sorting/storing/ dispatching tasks, high flexibility/versatility equipment shall be integrated only when and where balancing of productivity, effectiveness, efficiency and reliability leads to competitive benefits. Indeed, technology-driven options could easily lead to totally unmanned solutions, with autonomous rigs operating under remote overseeing. However, the all would result to deceptive outcomes and non justifiable waste. Through modelling and simulation, the assessment knowledge is built up, and this provides the decision support to develop and to manage properly tailored facilities.
The concept ideation, thereafter, shall duly establish the function assessment, based on the actual availability of facilities and technologies. For an example analysis, the study considers the regional mail delivery, with combined requirements: to grant the public service of covering the whole territory, with neighbourhood nets and widespread organisation; to guarantee accredited quality standards, as for due-time and reliability, within the balancing of universal service revenue and the value added dispatching. The latter fact means a standard sort centre shall address a huge mix of items: - letters, - flats, - packets, - parcels, which further distinguish by types: - surface, - express, - registered, - insured. The inward flow distinguishes four options: # street boxes, # post offices, # large posters, # sort centres; the outward flow addresses: ∘ final distribution, ∘ forwarding to other sort centres.
The function requirements analysis shows peculiarities that characterise each inward flow: # from the street boxes, letters and flats are mixed into sacks and require stamp cancelling; # from the post offices, bundles of flats and letters are gathered (to help trays forming) and forwarded with packets and parcels already stamped; # from the large posters, items arrive into sacks or containers, mostly requiring volume and franking checks; # from other sort centres, the mail comes in suitably separated into sets of homogeneous bunches or series of items. All in all, the mail feeding requires relevant pre-processing: in the first case, to be fully accomplished at the considered sort centre; possibly, anticipated by previous actions, on condition to propagate the pre-set selection during the dispatching, so that the ordered items and the joined information travel together. As for the outward flow, the single mail processing aims at separating the items to be allotted for local distribution, from the ones to be forwarded to the sort centres of different district areas. Of course, requirements analysis can lead to several set-ups. Typically, a postal code specifies a district, and a sort centre feeds more than one set of districts; a sorting facility has a fixed number of parallel outputs, and the overall incumbent requires to iterate the processing to cover the full set of considered districts. The forwarding to downstream sort centres, also, requires proper selection, depending on the number of regional links to feed (thus, on the outputs to be allocated); a linked up structure can be devised, distributing the items over a properly wide neighbourhood, and transferring the entire left-out mail to an assigned nearby sort centre; the connected loop will cover the full regional mail delivery, possibly, after assignments updating, with account of current work-on-progress.

Facility functional analysis.

The analysis shows that the transition from manual, to automatic processing completely covers the data flow and overall visibility is acquired to the regional boundary; thus, the corresponding economy of scope features are enabled, also if mixed automation solutions, with front-ends operators, are eventually selected. To that purpose, the processing facility could highly profit by functional modularity, allowing up-grading by fixture change or resource addition, still keeping the lay-out of the originally selected sort centre; this improvement can continuously be checked through modelling and simulation by virtual testing of the alternative lay-outs and schedules, thus, making easy the up-dating of the linked up structure, by pertinent assignments.
The facility analysis, thereafter, will readily progress, listing the existing technological aids, by respect to the duty-cycles required by any individual sort centre. A bird-eye view of the processing brings to acknowledge the following sections: ⌂ unload docks; ⌂ segregation and storage area; ⌂ manual preparation and, for items from street boxes, culling, facing and cancelling machine; ⌂ revenue protection, for items from large posters; ⌂ buffering area and feeding/singling forwarding by manual in-feed or by automatic in-feed; ⌂ optical coding recognition or video-coding support rig; ⌂ carrier and sorter machine; ⌂ amalgamation and loading area, for outward flow fitting. The last section represents the interface for further processing and exploits segregation and storage areas, pick-and-place handlers and roll containers, to prepare sacks, bundles or trays for the next duties. The second last section deserves additional specifications, to distinguish among the allotted items and duty phases: ▫ letter coding and pre-sorting or (given set of outputs) letter sorting duty; ▫ compact flat sorting duty; ▫ bundles and bulky items sorting duty; ▫ parcels sorting duty; ▫ registered, insured and express items processing area; ▫ manual sorting area. One needs notice that the special items follow reserved tracks to grant proper backtracking, security levels and allotted priorities; moreover, parallel manual tracks are used for un-

expected occurrences or at emergencies.
The process bird-eye view shows that: # productivity is mostly affected by the carrier and sorter machine; # effectiveness shall mainly depends on the inward and outward flows allocation; # efficiency will follow from the sort centre duty planning; # reliability comes out by managing servicing tasks and flexibility options. The centrality is taken up by code recognition and automatic sorting. The latter provides autonomy to the material flow; the former creates value added, by enabling data handling, vaulting and communication options. Now, machine vision and characters acknowledgement have reached proper levels of accuracy and standardisation, and the parallel track through the video-coding support rig is involved on rare occurrences, assuring manual machine-coding with full consistency for subsequent data flow propagation. Around these two sections, the pre-processing and post-processing complements are devised and implemented, to balance the performance figures with due account of return on investment. Of course, these figures are only partially evident, when dealing with an individual sorting area, when the study shell consider the regional mail delivery lay-out, on more complex linked up structures. Anyway, the sample discussions that follow already provide explanatory hints of the development methods.

Model definition and simulation setting.

The example sort centre includes all the previously mentioned resources, and the model definition takes into account the different functions, from mail collection and inward flow, to segregation tasks and outward flow. The actions to be undertaken depend on the arrival from the collecting points, and from the mail types. The ordinary items, further, distinguish in two classes: class A mail needs be delivered within the 24 hours, while class B mail could be delayed in case of high burden. The timing is specified, assuming a reference threshold for accepting the daily inward flow, say, 6 pm of every working day; then, later arriving items are ascribed to the next job order. A conventional work-day, accordingly, starts at 6 pm, and develops on four shifts, ending at 12 pm, 6 am, 12 am and 6 pm, respectively. The work-force could vary, with, generally, a peak at the first shift, when the class A mail needs be completely be processed.
The **Fig. 2** gives a synthetic view of the described functional arrangement, with indication of the work-stations and material paths. With **WITNESS** software, graphical representation of the lay-out helps providing intuitive hints on how the physical entities interact. The functions can, mainly, be gathered into four groups, say: # unloading, segregation and storage areas, pre-processing the different arrivals, to grant the inward flow homogeneity; # feeding/singling and sorting areas, to fulfil the "taking-in-charge" by address recognition and delivery allotment; # amalgamation and handling areas, to post-process the singled-out flows, with setting of the collections into convenient sacks or containers; # dispatching and delivery areas, to enable outward flows, with optimised management of the distribution and transport requirements.

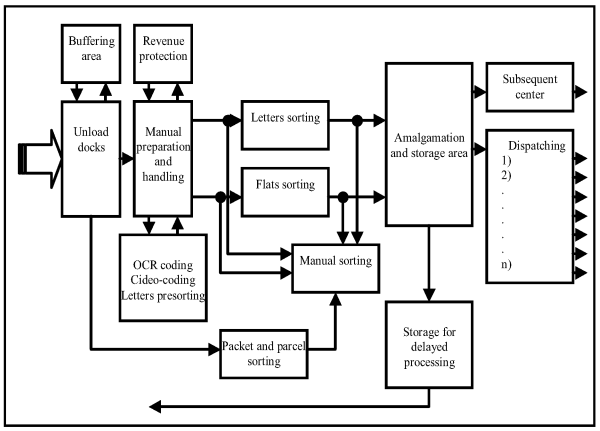


Fig. 2. Block-schema of an example sort centre.

Simulation environment description.

Referring to the recalled synthetic view, the second group functions are critical for the process automation: as recalled, innovative set-ups are explored, distinguishing the preliminary preparation, with resort to:

- the *culling*, *facing* and *cancelling*, CFC, machine, to distinguish the inward flow (coming from the mail-boxes), into: letters, letters with anomalies, flats and mixed (packets, etc.);
- the *letters coding* and *pre-sorting*, LCP, machine, to imprint the bar code and rough preliminary divide;
- the *revenue protection*, RP, machine, the verify the inward flow from large posters;

from the fully automatic processing of the items recognised to possess the requirements to be handled by:

- the *letters sorting*, LS, machine, to entirely divide the letters into the required distribution deliveries;
- the *compact flats sorting*, CLS, machine, to autonomously operate on that kind of deliveries;
- the *bundles* and *bulky* items *sorting*, BBS, machine, to process the more intrusive dispatching;
- the *parcels sorting*, PS, machine, to help processing larger weigh and/or size objects.

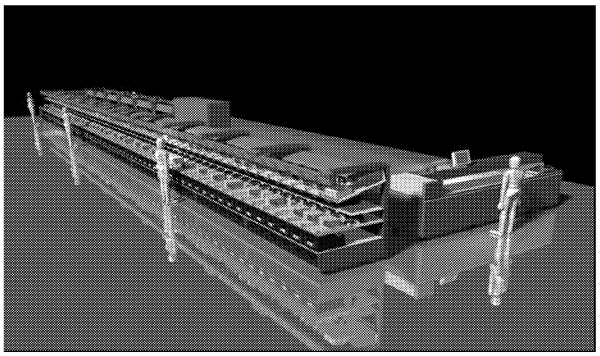


Fig. 3 Letter sorting machine.

The automatic flows require proper operators' attendance (see, e.g., the LS machine arrangement, **Fig. 3**), and in parallel the manual processing applies for items recognised to have anomalies. The preparation stages are, typically, required for the alien inflow (from

mail boxes and large posters), while the inmost flow (from post offices or sort centres) is already suitably arranged. This pre-setting aims at choosing by size (culling), at up-turning and stamp checking/cancelling, at imprinting the distinguishing mark, and at rough division into 30-40 streams, to optimise the following operations. Thereafter, the automatic sorting is performed (on parallel lines) only on properly acknowledged (and marked) items. The *letters sorting* and the *compact flats sorting* machines, **Fig. 3**, reach high performance, assuring sharp division into several tenths of allotments, so that the output dispatching can be programmed for several thousands deliveries, by iterated runs, timely empting the assigned sweepers. The bigger (and bulky) items require less effective singling and more complex handling actions, and the related machines exploit conveyors, with multiple branching off. Finally, the joint manual sorting section processes the not acknowledged (due to anomalies, etc) or the out-of standard items, or the extra flows (exceeding the facility capacity).

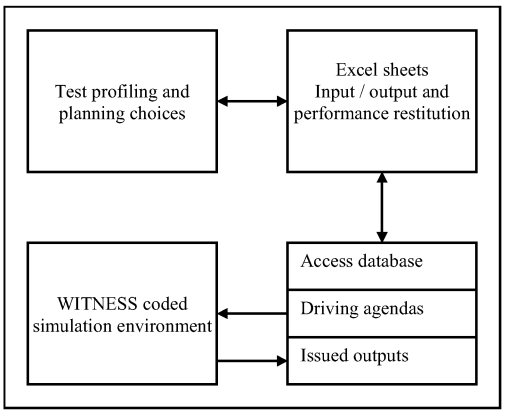


Fig. 4. Arrangement of the simulation environment.

The modelled sort centre, further to the physical entities, includes a set of related logical entities, specified to characterise the facility behavioural properties. Basically, one needs distinguish: ⊟ the governing rules, with specification at two levels: ▪ front-end commands, and ▪ remote supervisor; ⊟ the scheduled agendas, based on *typical* inward flows, provided by experimental observations. The resulting simulator duplicates the plant time behaviours, using the test agendas to steer the occurrence-driven evolution. At every moment, situation checks supply the process-data: input/output flows, work-on-progress, intermediate buffering, etc., for each local resource; cumulative data and history trends are, as well, available. These information play the role of process-descriptors, and are used to explore the actual facility performance, under the given driving agendas and governing strategies. To simplify comparisons, the descriptors are further processed, to supply explicit account of the overall performance, by means of synthetic facility indicators, namely, the above mentioned four figures: # productivity, # effectiveness, # efficiency, # reliability, shortly assessed in the next point.

The simulation environment organises according to four blocks, **Fig. 4**: - the processor, built with **WITNESS** software; - the relational data base, to collect the driving agendas and the issued outputs; - the Excel sheets interface, for outputs visualisation and performance analyses; - the pertinent text-files, to identify the testing progression and resource planning choices. The decision logic avails itself of the **WITNESS** software aid; e.g., for the *letters sorting* machine, the information flow (the below **Fig. 5**) follows assigned branching off cycles, started by the driving agenda. The control addresses sets of *elements*, blocks of information providing series of *actions* to be accomplished whenever the driving *occurrences* establish in the agenda, according to the enabling *rules*. Typical elements are: - *part*, entity, with identifying attributes, with annexed actions and rules, and situation (active/passive) label; - *machine*, operation resource, with specified work-cycle and duty status (including failure), and inward/outward rules; - *buffer*, storage resource; - *labour*, front-end human resource. The elements could assemble into modules, to help establishing meta-models, by aggregating the elemental functions, into meaningful macro-operators.

Simulation results analyses.

The simulation runs provide behavioural views of the facility as a all, and as individual resource, making the occurrence-driven dynamics to duplicate the time evolution of every details. Appropriate synthetic indicators are, further, computed the directly rank the facility (and agenda) performance.

The *productivity* aims at acknowledging the actually delivered amount of processed items on a one day term, or, separately, on the four daily shifts 6-12 pm, 0-6 am, 6-12 am and 12-6 pm; the figure splits to find out the average processing time, τ_κ, for the different handled pieces: κ, letters, flats, packets, parcels for ordinary (A or B class), registered, insured, express items. The *effectiveness* allows to know the ratio between consignees within the due-date and actually inward flow: $e = Q_S/Q_T = \Sigma D_j/\Sigma C_i$, the figure is usefully expressed separating the individual delivery of sorted items, D_j, and the disaggregated collection of mail, C_i, (arrived before 6 pm); then: $Q_R = Q_T - Q_S$, is the mail reminder to be shifted for next-day processing, remaining in the accumulation areas A_T. The *efficiency* is used to describe the utilisation ratio of each individual resource of the processing facility, and can be expressed extracting the run and idle periods of the equipment; alternatively, the figure is monetised, in keeping with human and machine fixed costs, c_H+c_M, by respect to direct processing costs, c_T, which includes amounts proportional to Q_T (or to Q_S); the balance is affected by investment locking up, with concern also of the A_P processing areas (for loading/unloading, handling, transportation, buffering, etc.). The *reliability* is assessed to specify the ability to keep the previous figures within given bounds; standard indices (MTBF, mean time between failures and MTTR, mean time to repair) are common reference, but the careful management of the (strategic, tactical, execution) flexibility helps improving the achievements, by adaptive job allocation, on-process dispatching, sort plan re-setting and up-dating, etc..

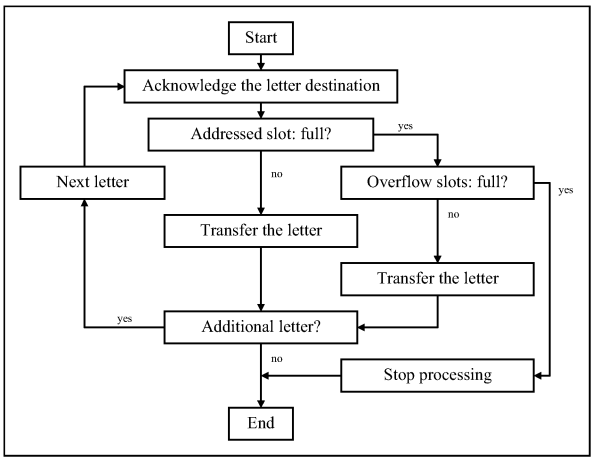


Fig. 5. Typical multiple-loop sorting cycle.

Sample investigations and checks.

An example can better explain a simulation run. Dealing, e.g., with a typical *sorting* machine, we look at the elements: feeder, sorter, stacker, sweeper and labour. The stacker, actually, corresponds to a 4x*n* matrix; the four arrays specify: the sorting slot, the filling ratio, the number of sorted items, the stack dispatching label; the columns: n, number of slots. The multiple-cycle sorting, **Fig. 5**, once loaded the mail tray by the feeder, at first assigns the slots to a subset of dispatching delivery, reserving a few slots for out-of-subset delivery, for downstream centres and for assigned-slot overflow, and starts transferring the items, until when one slot is full; then, switching the sweeping alarm on, tries diverting the exceeding items to the overflow slots, and, as these also are filled, the machine stops, until the sweeper operates; the cycle repeats, to process the all loaded tray. Then, a new sorting tasks is enabled, with different subset of dispatching delivery. The sweeper takes the items from the assigned delivery slots, and forwards the ordered bundles to the amalgamation and storage areas, to initialise the acknowledged dispatching.

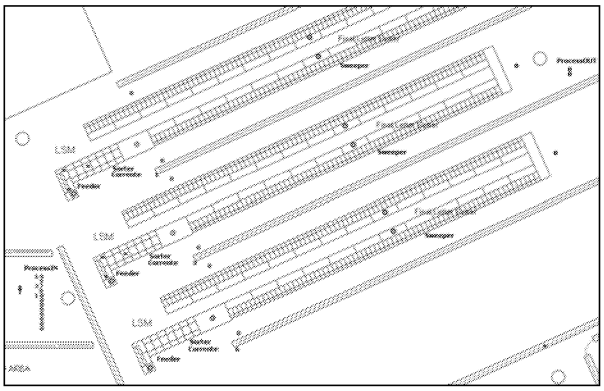


Fig. 6. Typical display: picture of paralleling three letter sorting machines.

Every simulation run directly interacts with the relational database. This organises by files, connected to each element, so to record the time history, and to compute cumulative and average indices. Actually, outputs are gathered for modules, accessed through proper masks, and displayed once reckoned the synthetic indicators. The programming environment avails of intuitive graphic displays, **Fig. 6**, and provides the time-histories of every resources, e.g.: the **Fig. 7**, shows the delivery of a letters coding an pre-sorting, LCPS, machine starting at 6 pm, and showing the instantaneous and the averaged flows (pcs/h) of the current day; the **Fig. 8**, provides the sweeping duty trends, due to filling up of overflow slots. Both duties exploit on-process operators (for the coding of item refused by the OCR, and for removing the trays at the filled slots), and the outputs correspond to example driving agendas; the involvement of front-end personnel requires careful analysis; the **Fig. 9**, e.g., the sequences of manual jobs (instantaneous and averaged number of actions), actually, accomplished along with the preparation, to feed the (parallel) letters and compact-flats sorting machines. The relational database and the Excel sheets interface similarly provide the on-progress the performance synthetic indicators.

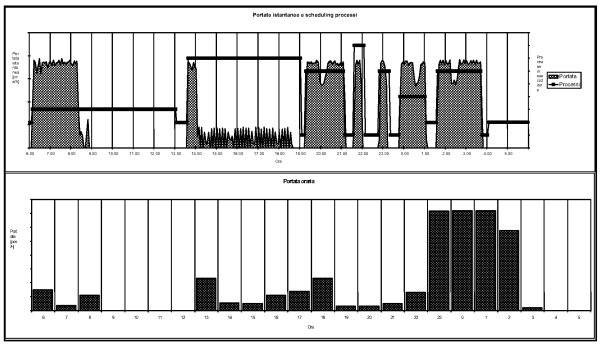


Fig. 7. Typical time-history of a letters coding and pre-sorting machine.

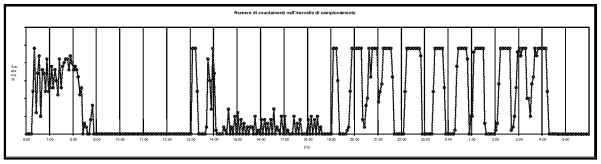


Fig. 8. Typical trends of overflow alarm and slot sweeping.

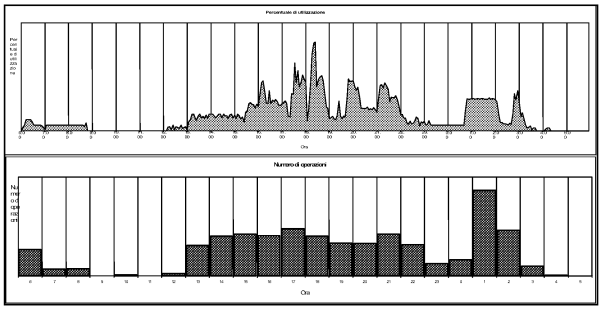


Fig. 9 . Typical manual jobs trends: instantaneous and averaged actions.

CONCLUDING COMMENTS.

The presentation discusses a noteworthy lifecycle design problem, to summarise benefits of modelling and simulation. These procedures are exploited during the conceptualisation steps, to assess the basic behavioural properties: on-duty performance,

effectiveness, reliability, maintenance, adaptivity, expansibility, .., by using virtual prototyping to acknowledge the on-duty performance within actual running conditions. The example case provides explanatory hints, to turn the value chain toward product-service supply; the management of the on-process (strategic, tactical, execution) flexibility is possible if bottom-up information loops close, to modify the resource planning, to select alternative operation schedules, to resort to additional resources and technologies, etc., making possible on-progress improvements, as the need arises. The presentation stresses on the design steps: requirements analysis; function specification; model development; behaviour simulation; at the same time the on-duty characteristics are properly recalled: resource planning; operation management; performance assessment; facility up-grading; this allows proper expansion of mixed-automation set-ups, as soon the performance checks show actual return on investment.

The explanatory example bears special interest, to justify the transition from manual to automatic processing, helping to assess the extent of mixed-mode automation by a careful preliminary investigation on the actual achievements. This transition is assessed on three opportunities: # joint flow of mail items and data, in order to provide process visibility and information added value; # functional modularity and up-grading, by fixture change or resource addition; # modelling and simulation support for virtual testing of lay-outs and schedules. The three of them are tightly interconnected, and represent the new paradigms of competitiveness, created by lifecycle design and product-process supply chains.

The discussion offers a noteworthy example case, showing how modelling, simulation and virtual testing are emerging key technologies to support scope-driven design; they are critical at the point of service to improve the on-duty effectiveness and to assess the economical return of competing lay-outs. The actual value of the approach is in its ability to capture and to represent knowledge, so that confident predictions are achieved for on-duty running conditions. This way, the test/evaluate/modify phase is faced as initial design step, providing **[18]**:

- instant, clear, accurate visibility into the status and performance of actual on-duty set-ups;
- quick confrontation of competing lay-outs behaviours and/or process planning agendas;
- rapid assessment of how resource up-grading or down-grading affects operation productivity;
- fast and reliable exploration of transient effects, unexpected occurrences and disturbances.

It is important to note that there is a wide range distinguishing the current state of practice, from the existing state of art. Many of the design approaches are acknowledged by leading-edge companies, still not exploited in the daily routine. An important spur to make modelling, simulation and virtual testing rule, not exception, would be integrating mixed-automation processes, where on-line operators are interfaced with robotic rigs, and addressing real-time controllers, where governing modules could equivalently connect to real either to simulated processes. The derivation of mixed causal- and heuristic-framed models allows linking algorithmic and procedural blocks and emulating human actions with coherent progression of material and information flows.

REFERENCES.

[1] G.M. Acaccia, R.C. Michelini, R.M. Molfino: "Design of intelligent governors for the material-handling equipment of automated factories", T.Sata, G.Olling, Eds.: 'Software for Factory Automation', North-Holland, Amsterdam, 1989, pp. 297-312.

[2] G.M. Acaccia, R.C. Michelini, R.M. Molfino: "Govern-for-flexibility of tool-stock and delivery operation for factory automation", G.J. Olling, Z. Deng Eds: 'Information Technology for Advanced Manufacturing Systems', Elsevier Sci. Pub. Co., 1991, pp. 483-490.

[3] G.M. Acaccia, R.C. Michelini, R.M. Molfino: "Computer-intelligence in flexible manufacturing", J.L. Alty, L.I. Mikulich Eds.: Industrial Applications of Artificial Intelligence, North-Holland, 1992, pp. 344-349.

[4] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino: "An expert simulator for factory automation", G.J.Olling, F.Kimura Eds: Human Aspects in Computer-Integrated Factories, North-Holland Pub., 1992, pp. 797-804.

[5] G.M. Acaccia, M. Callegari, R.C. Michelini, R.M. Molfino, P. Peri, A. Ricci: "Control automation of a multioperational section for the flexible manufacturing of highly-diversified products", I. Mezgar, P. Bertok Eds. 'Knowledge Based Hybrid Systems', North-Holland, 1993, pp. 193-202.

[6] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino: "Flexible manufacturing with integrated control and management", M.B. Zaremba, B. Prasad Eds.: 'Modern Manufacturing: Control & Technology', Springer Verlag, London, 1994, pp. 225-253.

[7] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino, R.P. Razzoli: "Shop controller-and-manager for intelligent manufacture", S. Tzafestas Ed.: 'Control and Management of Manufacturing Systems', Springer Verlag, London, 1997, pp. 219-254.

[8] R.C. Michelini, G.L. Kovàcs: "Knowledge setting and govern-for-flexibility in lean manufacturing", A.B. Baskin, G.L. Kovàcs, G. Jacucci, Eds: 'Co-operative Knowledge Processing for Engineering Design', Kluwer Acad. Pub., Norwell, 1999, pp. 61-82.

[9] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino, R. Razzoli: "Artefact integration by concurrent enterprises and productive break-up", G.Jacucci, G.J.Olling, K.Preiss, M.Wozny, Eds.: 'Globalisation of Manufacturing in the Digital Era', Kluwer Acad. Pub., Boston, 1999, pp. 221-234.

[10] Lanner Group: http://www.lanner.com.

[11] G.M. Acaccia, M. Conte, D. Maina, R.C. Michelini, R.M. Molfino: "Integrated manufacture of high-standing dresses for customised satisfaction", G. Jacucci, G.J. Olling, K. Preiss, M. Wozny, Eds.: 'Globalisation of Manufacturing in the Digital Era',

Kluwer Academic Pub., Boston, 1999, pp. 511-523.
[12] R.M.Molfino, A. Lacchini, G. Maggiolo, R.C. Michelini, R.P. Razzoli: "Re-engineering issues in automatic assembly", G. Jacucci, G.J. Olling, K. Preiss, M.Wozny, Eds.: 'Globalisation of Manufacturing in the Digital Era', Kluwer Acad., Boston, 1999, pp. 603-616.
[13] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino, R.P. Razzoli: "Computer integrated assembly for cost effective developments", C.T. Leondes Ed.: Computer Integrated Manufacturing, Vol. II, CRC Press LLC, Boca Raton Fl, 2001, pp. 2.01-2.68.
[14] R.C. Michelini, G.M. Acaccia, M. Callegari, R.M. Molfino, R.P. Razzoli: "Robots design with application to manufacturing", C.T. Leondes Ed.: Robotics and Artificial Intelligence in Manufacturing, Vol. VII, CRC Press LLC, Boca Raton, 2001, pp. 7.01-7.62.
[15] R.C. Michelini, G.M. Acaccia, M. Corvi, W. Ferebauer: "Assisted mails sorting and forwarding stand: performance analysis & ergonomic assessment", XI Intl. Prolamat Conf. Digital Enterprise, G.L Kovacs, P. Bertok, G. Haidegger, Budapest, 2001, pp. 543-554.
[16] http://www.logistics-assembly.siemens.com
[17] http://www.elsag.it
[18] IMTI: The Integrated Manufacturing Technology Initiative http://www.IMTI21.org

Stable and robust feedback linearization adaptive fuzzy control of SISO nonlinear systems using Takagi Sugeno fuzzy models

D. BEN HALIMA and M. CHTOUROU

Research Unit on Intelligent Control, design and Optimization of complex Systems (ICOS)
Departement of Electrical Engineers
National School of Engineers of Sfax, P.B.W, 3038 Sfax, Tunisia
Tel: (+216) 74 274088 Fax (+216) 74 275595

e-mail: donia.benhalima@voila.fr mohamed.chtourou@enis.rnu.tn

Abstract

An adaptive fuzzy controller is constructed from a collection of fuzzy IF-THEN rules whose parameters are adapted according to adaptive laws for the purpose of controlling a nonlinear plant to track a reference trajectory. The adaptive fuzzy controller consists of two parts. The first one is called certainty equivalent fuzzy controller. It's an indirect adaptive fuzzy controller witch is constructed using fuzzy plant dynamic approximators. The second one is a direct fuzzy robutstifying term witch is added to the former controller to ensure system stability and robustness. The updating laws of fuzzy systems parameters are obtained using Lyapunov stability approach. Simulation results show the effectiveness of proposed controller in tracking of the inverse pendulum output.

Keywords: *Adaptive fuzzy control, Certainty equivalent control, Supervisory control, Robustness, Stability.*

1. Introduction

Adaptive control theory has received a significant research attention in the past few years and has evolved as a powerful methodology for designing feedback controller for nonlinear systems. However, this control technique relies on exact mathematical model of nonlinear systems to obtain linear input output behaviour [1-4]. For many nonlinear dynamical systems, it is difficult to develop an accurate mathematical model for them.

Fuzzy system and conventional control are complementary rather than competitive. Fuzzy system can borrow and refer from conventional control to consolidate into stronger area. The integration of fuzzy systems and conventional control has recently found extensive applications for a wide of variety of the system control. Because, fuzzy system need no accurate mathematical model of the system to control. Based on the universal approximation theorem [5], many researchers presented indirect or direct adaptive fuzzy control schemes for nonlinear systems [6-16]. For specifically, direct adaptive fuzzy controllers use fuzzy systems as controllers, therefore linguistic fuzzy control rules can be directly incorporated into the controllers. On the other hand, indirect adaptive fuzzy controllers use fuzzy systems to model the plant and construct the controllers assuming that the fuzzy systems represent (approximately) the true plant; therefore fuzzy IF-THEN rules describing the plant can be directly incorporated into the indirect adaptive fuzzy controller. For both of them, fuzzy parameters would be adjusted on-line by any adaptive laws such as optimization technique which lack for analytical results about stability and performance. To overcome these problems, several elegant adaptive fuzzy control approaches have been proposed based on Lyapunov's stability theory [7-10, 14-17]. Wang [15], has proposed parameter adaptation laws based on Lyapunov synthesis and he has introduced a supervisory control term to guarantee stability and robustness of the closed loop system. But, bounds on the unknown plant dynamics must be known. Generally, this assumption requires an exact model of the plant. To solve this problem, Chen et al [13] are inspired of H^{∞} technique and they have added a robust controller term to an indirect adaptive fuzzy controller. However, their controller results in fuzzy certainty equivalent controller influence attenuation because it's a large gain. Kang et al [18] have proved this disadvantage, and they have proposed a solution for this problem. Park et al [12] have proposed adaptive control law consisting of two terms. One is the certainty equivalent control and the other is similar to supervisory control in [15] or to a robust one in [13] and it is a bounding control. They assume that the reconstructions errors satisfy certain partially known bounding conditions that are, in practical, not easy to know.

The purpose of this paper is to develop a stable and robust adaptive algorithm using fuzzy systems for single-input single output (SISO) unknown nonlinear dynamical systems. The algorithm is inspired from Wang's one [15]. The control law consists of two parts. The first one is the indirect adaptive control called certainty equivalent control law. The second is an adaptive fuzzy control used to guarantee stability and robustness of adaptive control law. Parameter adaptation laws of the second control term and of unknown nonlinearities are obtained using Lyapunov's stability approach.

This paper is organised as follows: Section 2 presents the problem statement and introduces Wang's adaptive fuzzy controller [15]. The proposed adaptive fuzzy algorithm and the stability analysis of the closed loop system are

presented in section 3. In section 4, simulation results show that the proposed controller guarantees the tracking error convergence in the small neighbourhood of zero. Conclusion is presented in section 5.

2. Problem Statement

Consider the nth-order nonlinear system described by the following equation:

$$y^{\cdot} = f(x) + g(x)\ u \qquad (1)$$

where $y \in \Re$ is the measured output, $u \in \Re$ is the control input, $x = \left[\ y, \dot{y}, \ldots, y^{(n-1)}\ \right]^T$ is the state vector, $y^{(i)}$ is the i-th time derivative of the output y, and f and g are two unknown nonlinear functions $\left(\Re^n \longrightarrow \Re\right)$, n denotes the system order.

Then we represent system (1) in the state space:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_n &= f(x_1, x_2, ..., x_n) + g(x_1, x_2, ..., x_n)\, u \\ y &= x_1 \end{aligned} \qquad (2)$$

The state vector of the system is assumed to be available for measurement.

The control objective is to determine a feedback control $u = u(x, \theta)$ and an adaptation law for adjusting the vector parameter θ such that [15]-[16]:

i) The closed loop system must be globally stable and robust in the sense that all variables $x(t)$, $\theta(t)$ and $u(x, \theta)$ must be uniformly bounded.

ii) The system output tracks a given desired trajectory $y_m(t)$, with an acceptable accuracy, i.e. the tracking error $e = y_m - y$ should be as small as possible under the constraints defined in i)

It is assumed that the desired trajectory $y_m(t)$, and it's time derivatives $y_m^{(1)}(t)$, $y_m^{(2)}(t)$, ...and $y_m^{(n)}(t)$ are smooth and bounded.

If the functions f and g are known, then the control law is giving by [15]:

$$u = \frac{1}{g(x)}\left[-f(x) + y_m^{(n)} + k^T E\right] \qquad (3)$$

where $E = \left[e, \dot{e}, ..., e^{(n-1)}\right]^T$ and $k = \left[k_n, ..., k_1\right]^T \in \Re^n$ be such that all roots of the polynomial $s^n + k_1 s^{n-1} + ... + k_n$ have negative real parts.

Substuting (3) in (2) results in:

$$e^{(n)} + k_1 e^{(n-1)} + ... + k_n e = 0 \qquad (4)$$

which implies that:

$\lim_{t \longrightarrow \infty} e(t) = 0$.

If functions f and g are unknown, they will be replaced by fuzzy systems $\hat{f}(x|\theta_f)$ and $\hat{g}(x|\theta_g)$. Where θ_f and θ_g denote the fuzzy systems parameters defining $\hat{f}$ and $\hat{g}$, respectively.

Giving the estimates $\hat{f}\ (\ x\ |\ \theta_f\ )$ and $\hat{g}\ (x\ |\ \theta_g)$, the resulting control law becomes:

$$u_c = \frac{1}{\hat{g}(x, \theta_g)}\left[-\hat{f}(x, \theta_f) + y_m^{(n)} + k^T E\right] \qquad (5)$$

for witch u_c is called the certainty equivalent controller. Applying (5) to (2), following equation will be obtained:

$$\begin{aligned} \dot{x}_n &= f(x_1, ..., x_n) + g(x_1, ..., x_n)\ u_c \\ y_m^{(n)} - \dot{x}_n &= y_m^{(n)} - f(x_1, ..., x_n) - g(x_1, ..., x_n)\ u_c + k^T E \\ &\quad -k^T E\ + \hat{f}(x|\theta) - \hat{f}(x|\theta) \end{aligned}$$

$$e^{(n)} = -k^T E + \left[\hat{f}(x|\theta_f) - f(x)\right] + \left[\hat{g}(x|\theta_f) - g(x)\right] u_c \qquad (6)$$

or equivalently

$$\dot{E} = A_c E + b_c \left\{\left[\hat{f}(x|\theta_f) - f(x)\right] + \left[\hat{g}(x|\theta_f) - g(x)\right] u_c\right\} \qquad (7)$$

Where

$$A_c = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ -k_n & -k_{n-1} & \ldots & \ldots & \ldots & -k_1 \end{bmatrix} \quad b_c = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \ldots \\ 1 \end{bmatrix}$$

It's obvious that A_c is a hurwitz matrix. Therefore, for a $(n \times n)$ symmetric positive matrix Q, there exists a unique positive definite symmetric $(n \times n)$ matrix P that satisfies the Lyapunov equation:

$$A_c^T P + P A_c = -Q \qquad (8)$$

Let V_e be a Lyapunov's function defined as:

$V_e = \frac{1}{2} E^T P E$.

The time derivative of V_e is:

$$\begin{aligned} \dot{V}_e &= \frac{1}{2}\dot{E}^T P E + \frac{1}{2} E^T P \dot{E} \\ &= -\frac{1}{2} E^T Q E + E^T P b_c \times \\ &\left[\left[\hat{f}(x|\theta_f) - f(x)\right] + \left[\hat{g}(x|\theta_g) - g(x)\right] u_c\right] \end{aligned}$$

To guarantee the boundedness of e and x, we must require that $\dot{V}_e \le 0$, when V_e is greater than a large constant $\overline{V}$ (which is a constant specified by the designer). In this

context, we append another control term u_s as to the u_c. Then the final control law is [15]:

$$u = u_c + u_s \tag{9}$$

The additional robustifying control is called a supervisory control [15]. To determine u_s, we suppose that $0 < g_l(x) \le g(x) \le g_u(x) < \infty$ and

$$\left| f(x) \right| \le f_u(x) < \infty$$

Applying (9) to (2), we obtain:

$$\begin{aligned}
&\dot{x}_n = f(x_1,\ldots,x_n) + g(x_1,\ldots,x_n)\left[u_c + u_s\right] \\
&y_m^{(n)} - \dot{x}_n = y_m^{(n)} - f(x_1,\ldots,x_n) - g(x_1,\ldots,x_n)\left[u_c + u_s\right] \\
&\quad + k^T E - k^T E + \hat{f}(x|\theta) - \hat{f}(x|\theta)
\end{aligned}$$

$$\dot{E} = A_c E + b_c \left\{ \left[\hat{f}(x|\theta) - f(x)\right] + \left[\hat{g}(x|\theta) - g(x)\right] u_c - g(x) u_s \right\} \tag{10}$$

Using (10) and (8), we have:

$$\begin{aligned}
&\dot{V}_e = -\frac{1}{2} E^T Q E + E^T P b_c \times \\
&\left\{ \left[\hat{f}(x|\theta_f) - f(x)\right] + \left[\hat{g}(x|\theta_g) - g(x)\right] u_c - g(x) u_s \right\} \\
&\le -\frac{1}{2} E^T Q E + \left| E^T P b_c \right| \times \\
&\left\{ \left|\hat{f}(x|\theta_f)\right| + \left| f(x) \right| + \left|\hat{g}(x|\theta_g) u_c\right| + \left|g(x) u_c\right| \right\} \\
&-E^T P b_c g(x) u_s \le -\frac{1}{2} E^T Q E + \left| E^T P b_c \right| \times \\
&\left\{ \left|\hat{f}(x|\theta_f)\right| + \left| f_u(x) \right| + \left|\hat{g}(x|\theta_g) u_c\right| + \left|g_u(x) u_c\right| \right\} \\
&-E^T P b_c g(x) u_s
\end{aligned}$$

if u_s is expressed by (11), we can guarantee that $\dot{V}_e \le 0$.

$$\begin{aligned}
u_s &= I_l^* \, sgn\left(E^T P b_c\right) \frac{1}{g_l(x)} \times \\
&\left[\left|\hat{f}(x|\theta_f)\right| + \left|f_u(x)\right| + \left|\hat{g}(x|\theta_g) u_c\right| + \left|g_u(x) u_c\right| \right]
\end{aligned} \tag{11}$$

where $I_l^* = \begin{cases} 1 & \text{if } V_e > \overline{V} \\ 0 & \text{if } V_e \le \overline{V} \end{cases}$

However, the additional term u_s has some disadvantages in the practical applications. In the one hand, bounding functions g_u, g_l, and f_u need to be known. In general, it not easy to determine these bounds for the unknown system. In the other hand, u_s has a high frequency chattering due to the existence of $I_l^* \, sgn\left(E^T P b_c\right)$ in its expression. To solve this problem, a fuzzy system is proposed to generate the supervisory control u_s.

3. Adaptive fuzzy controller

The next task, is to replace $u_s(E|\theta_s)$, $\hat{f}(x|\theta_f)$ and $\hat{g}(x|\theta_g)$ by specific formula of fuzzy systems and develop an adaptation law for adjusting the parameters in the fuzzy systems for the purpose of forcing the tracking error to converge to zero.

3.1. Takagi-Sugeno fuzzy systems

A multiple-input single-output (MISO) fuzzy system is a nonlinear mapping from an input vector $X = [x_1, x_2, \ldots, x_n]^T \in \Re^n$ to an output $y \in \Re$.

For a zero-order Takagi-Sugeno fuzzy systems, the fuzzy rule base contains a collection of fuzzy IF-THEN rules of the form:

$$\begin{aligned}
R^L: \quad &\text{IF}\left(x_1 \text{ is} F_1^L\right) \text{and} \ldots \text{and}\left(x_n \text{ is} F_n^L\right) \\
&\text{THEN } y = y^L
\end{aligned} \tag{12}$$

where F_i^L is the label of the fuzzy set corresponding to the variable x_i, for L=1, 2, …, n, and y^L is a constant parameter representing the consequent part of the fuzzy rule L.

In this paper, we use a set of fuzzy systems with singleton fuzzifier, and the product inference for the "and" operator. Consequently, the final output value of the fuzzy system can be expressed as:

$$y = \frac{\sum_{L=1}^{N} y^L \left(\prod_{i=1}^{n} \mu_{F_i^L}(x_i) \right)}{\sum_{L=1}^{N} \left(\prod_{i=1}^{n} \mu_{F_i^L}(x_i) \right)} = y^T \zeta(x) \tag{13}$$

where N is the fuzzy rule number and $\mu_{F_i^L}(x_i)$ is represented by a Gaussian membership function expressed by:

$$\mu_{F_i^L}(x_i) = \exp\left(-\frac{1}{2} \left(\frac{x_i - a_i^L}{\sigma_i^L} \right)^2 \right).$$

And $\zeta^L(x) = \dfrac{\prod_{i=1}^{n} \mu_{F_i^L}(x_i)}{\sum_{L=1}^{N} \left(\prod_{i=1}^{n} \mu_{F_i^L}(x_i) \right)}$; L = 1, 2, …, N.

y^L are adjustable parameters. It is obvious that the fuzzy system is linear of y^L.

3.2. Adaptation law

In this section, we replace $u_s(E|\theta_s)$, $\hat{f}(x|\theta_f)$ and $\hat{g}(x|\theta_g)$ by specific formula of fuzzy systems of (13). We need to adjust the parameters in the fuzzy systems for obtaining our purpose.

Let us:

$$\theta_f^* = \arg\min_{\theta_f} \left[\sup_x \left| \hat{f}(x|\theta_f) - f(x) \right| \right] \tag{14}$$

$$\theta_g^* = \arg\min_{\theta_g} \left[\sup_x \left| \hat{g}(x|\theta_f) - g(x) \right| \right] \tag{15}$$

$$\theta_s^* = \arg\min_{\theta_s} \left[\sup_E \left| \hat{u}_s(x|\theta_s) - u_s(E) \right| \right] \tag{16}$$

The $\hat{f}(x|\theta_f)$, $\hat{g}(x|\theta_g)$ and $u_s(E|\theta_s)$ are constructed as:

$$\hat{f}(x|\theta_f) = \theta_f^T \zeta(x) \quad (17)$$

$$\hat{g}(x|\theta_g) = \theta_g^T \zeta(x) \quad (18)$$

$$\hat{u}_s(E|\theta_s) = \theta_s^T \zeta(E) \quad (19)$$

Let $V(t) = V_c(t) + \frac{1}{2\gamma_1}\phi_f^T\phi_f + \frac{1}{2\gamma_2}\phi_g^T\phi_g + \frac{1}{2\gamma_3}\phi_s^T\phi_s$

If we choose the following adaptation laws to adjust fuzzy parameters, we can guarantee that $\dot{V}(t) < 0$ then the stability of the closed loop system we be assured.

$$\dot{\theta}_f = -\gamma_1 E^T P b_c \zeta(x) \quad (20)$$

$$\dot{\theta}_f = -\gamma_2 E^T P b_c \zeta(x)\ u_c \quad (21)$$

$$\dot{\theta}_s = \begin{cases} -\gamma_3 E^T P b_c \hat{g}(x|\theta_g)\zeta(E) & \text{if } E^T P b_c \hat{g}(x|\theta_g)\hat{u}_s(E|\theta_s) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

3.3. Design of the adaptive fuzzy controller

The adaptive fuzzy controller design can be summarized by the following steps.

Step 1: Determine the $k_1, k_2, \ldots, k_n$, and a positive definite matrix Q described in section 2 to obtain a Lyapunov's symmetric matrix P.

Step 2: Specify the constant constraint $\bar{V}$.

Step 3: Construct the initial controller. In the other word, construct the fuzzy rule basis for the fuzzy systems $\hat{f}(x|\theta_f)$, $\hat{g}(x|\theta_g)$ and $\hat{u}_s(E|\theta_s)$ as the forms:

$$R_f^L: \text{IF } (x_1 \text{ is } F_1^{L_1}) \text{ and } \ldots \text{and } (x_n \text{ is } F_n^{L_n}) \text{ THEN } \hat{f}(x|\theta_f) = T^L \quad (23)$$

$$R_g^L: \text{IF } (x_1 \text{ is } F_1^{L_1}) \text{ and } \ldots \text{and } (x_n \text{ is } F_n^{L_n}) \text{ THEN } \hat{g}(x|\theta_g) = Z^L \quad (24)$$

$$R_s^L: \text{IF } (e \text{ is } F_1^{L_1}) \text{ and} \ldots \text{and } (e^{(n)} \text{ is } F_n^{L_n}) \text{ THEN } \hat{u}_s(E|\theta_s) = H^L \quad (25)$$

where $L_i = 1, 2, \ldots, m_i$, $i = 1, 2, \ldots, n$, and T^L, Z^L and H^L are constant parameters.

Step 4: Determine the output of each fuzzy system $\hat{f}(x|\theta_f)$, $\hat{g}(x|\theta_g)$ and $\hat{u}_s(E|\theta_s)$ using, respectively, (17), (18) and (19).

Step 5: Apply the feedback control u (9) to the plant (2).

Step 6: Adjust the parameters of the fuzzy systems using equations (20) – (22).

4. Simulation results

In this section, we use the proposed adaptive fuzzy controller to control the inverted pendulum system (figure 1) to track some forms of trajectory. Let $x_1 = \theta$ and $x_2 = \dot{\theta}$, the dynamic equations of the inverted pendulum system are [15]:

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = \frac{g\sin x_1 - \frac{mlx_2^2 \cos x_1 \sin x_1}{m_c + m}}{l\left(\frac{4}{3} - \frac{m\cos^2 x_1}{m_c + m}\right)} + \frac{\frac{\cos x_1}{m_c + m}}{l\left(\frac{4}{3} - \frac{m\cos^2 x_1}{m_c + m}\right)} u \quad (26)$$

where g = 9.8 m/s^2 is the acceleration due to gravity, m_c =1 kg is the cart mass, m = 0.1 kg is the pole mass, l =0.5m is the half-length of the pole, and u is the applied force.

We choose a reference signal defined as follow:

$$y_m(t) = \frac{\pi}{30}\sin(t) \quad (27)$$

We choose $k_1 = 2$, $k_2 = 1$, and $Q = \begin{bmatrix} 10 & 5 \\ 5 & 10 \end{bmatrix}$. We obtain $P = \begin{bmatrix} 15 & 5 \\ 5 & 5 \end{bmatrix}$. We choose 5 membership functions for x_1 and 3 ones for x_2 covering the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Two fuzzy systems each of them is composed of fifteen rules are used to estimate f and g. To estimate the fuzzy supervisory control, 3 membership functions for e and $\dot{e}$ covering the interval [-0.5 0.5], are used.

To show the performance and robustness of the proposed controller, we simulate it for the last inverted pendulum tracking problem. The initial condition is x(0) = $[-0.05\ 0]^T$. The adaptation rate constants are setting to $\gamma_1 = 40$, $\gamma_2 = 4$, $\gamma_3 = 400$.

Figures 2 and 3 show the simulation results with $m_c(t)$ = 1 kg and $m_c = 2$ kg, for $t \geq 5$s, respectively. It is obvious that the system output tracks, sufficiently, the desired output.

In the purpose of comparison, we simulate the Wang's method for the same initial condition, the same adaptation rate constant, for the same values of m_c. f_u, g_u, $\bar{V}$ and g_l are setting, respectively, to 1.68, 1.46, 0.001 and 1.32. Figures 4 and 5 show the simulation results for Wang's method. Comparing figures 2 and 3 with figures 4 and 5 respectively, we can remark the degradation of tracking performance. We don't see any chattering occurs in figure 6 (a). But, a relatively high one occurs for the Wang's method.

5. Conclusion

In this paper, an adaptive fuzzy controller is developed. The proposed algorithm does not require an accurate mathematical model of the system under control. It's capable to guarantee the globally stability of the resulting closed loop systems in the sense that all signals involved are uniformly bounded. The control law consists of two terms the certainty equivalent controller and the robustifying one which is provided by a fuzzy system. Fuzzy parameters of the unknown system and of robustifing controller are updated based on Lyapunov stability theory. Proposed adaptive fuzzy controller is used to control the

inverted pendulum and the simulation results show that fuzzy controller could perform successful tracking.

References

[1] J. –J.E. Slotine and W. Li, "Applied Nonlinear Control", Prentice-Hall, Englewood Cliffs, NJ, 1991.

[2] S. Sastry and M. Bodson, "Adaptive control, Stability, convergence and robustness", Prentice-Hall, Englewood Cliffs, NJ, 1989

[3] A. Isidori, "Nonlinear control systems", Berlin: Springer-Verlag, 1989.

[4] P. A. Ioannou and J. Sun, "Robust Adaptive control", Prentice-Hall, Upper Saddle River, NJ, 1994.

[5] L. X. Wang, "Fuzzy systems are universal approximators", proceeding of IEEE Conference on Fuzzy Systems, San Diego, 1992, pp. 1163-1170.

[6] L. X. Wang, "Adaptive fuzzy systems and control", Prentice-Hall, Englewood Cliffs, NJ, 1994.

[7] L. X. Wang, " Stable adaptive fuzzy control of nonlinear systems", IEEE trans. Fuzzy Sys., vol. 1, 1993, pp. 146-155.

[8] M. Hojati and S. Gasor, "Hybrid adaptive identification and control of nonlinear systems", IEEE Trans. Fuzzy Syst., vol. 10, n. 2, 2002, pp. 198-210.

[9] J. Wang, A.B. Rad and P.T. Chan, "Indirect adaptive fuzzy sliding mode control: Part I: Fuzzy switching", Fuzzy Sets and syst. Vol. 122, 2001, pp. 21-30.

[10] S. Tong, T. Wang and T. Tong, "Fuzzy adaptive output tracking control of nonlinear systems", Sets and syst. Vol. 111, 2000, pp. 169-182.

[11] P.T. Chan, A.B. Rad and J. Wang, "Indirect adaptive fuzzy sliding mode control: Part II: Parameter projection and supervisory control", Fuzzy Sets and syst. Vol. 122, 2001, pp. 31-43.

[12] J. H. Park, S. J. Seo and G. T. Park, "Robust adaptive fuzzy controller for nonlinear system using estimation of bounds for approximation errors", Fuzzy Sets and syst. Vol. 133, 2003, pp. 19-36.

[13] B. S. Chen, C. H. Lee and Y. C. Chang, " H^{∞} tracking design of uncertain nonlinear SISO systems: adaptive fuzzy approach", IEEE trans. Fuzzy Sys., vol. 4, n. 1, 1996, pp. 32-43.

[14] H. N. Nounou and K. M. Passiono, "Stable auto-tuning of adaptive fuzzy/neural controllers of nonlinear discrete time systems", IEEE trans. Fuzzy Sys., vol. 12, n. 1, 2004, pp. 70-83.

[15] L. X. Wang, " Adaptive fuzzy controllers with application to inverted pendulum tracking", IEEE trans. SMC- Part B vol. 26, n. 5, 1996, pp. 677-691.

[16] P. S. Yoon, J. H. Park, and G. T. Park, " Adaptive Fuzzy Control of Nonaffine Nonlinear systems Using Takagi-Sugeno Fuzzy Models", IEEE Int. Conf. Fuzzy Syst. Vol. 2, Dec. 2001, pp. 642-645.

[17] Y. Gao and M. Joo Er, "Online adaptive fuzzy neural identification and control of a class of MIMO nonlinear systems", IEEE trans. Fuzzy Sys., vol. 11, n. 4, 2003, pp. 462-467.

[18] H.J. Kang, H. Lee and M. Park, "Comments on H^{∞} tracking design of uncertain nonlinear SISO systems: adaptive fuzzy approach", IEEE trans. Fuzzy Sys., vol. 6, n. 4, 1998, pp. 605-606

.

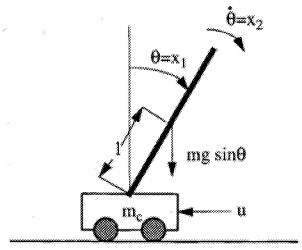
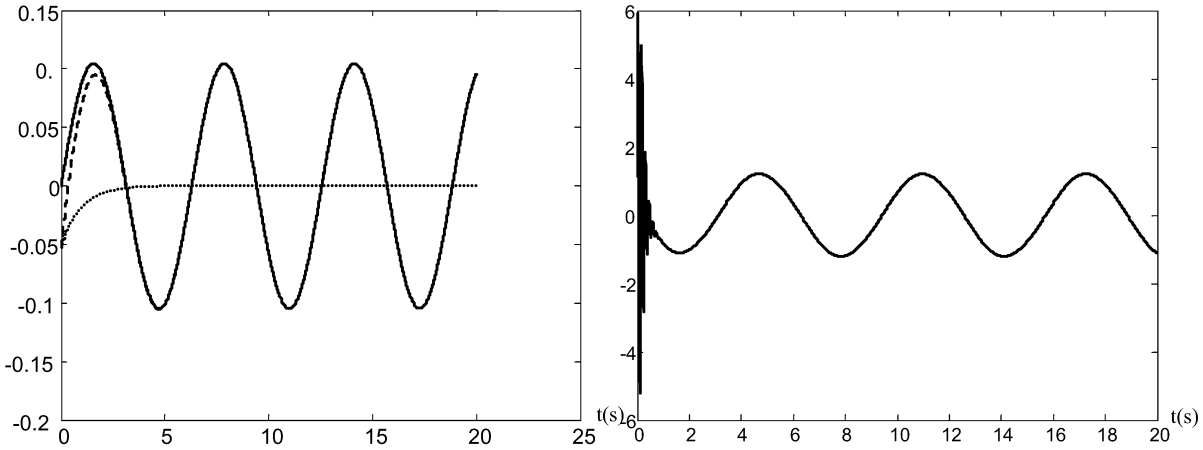
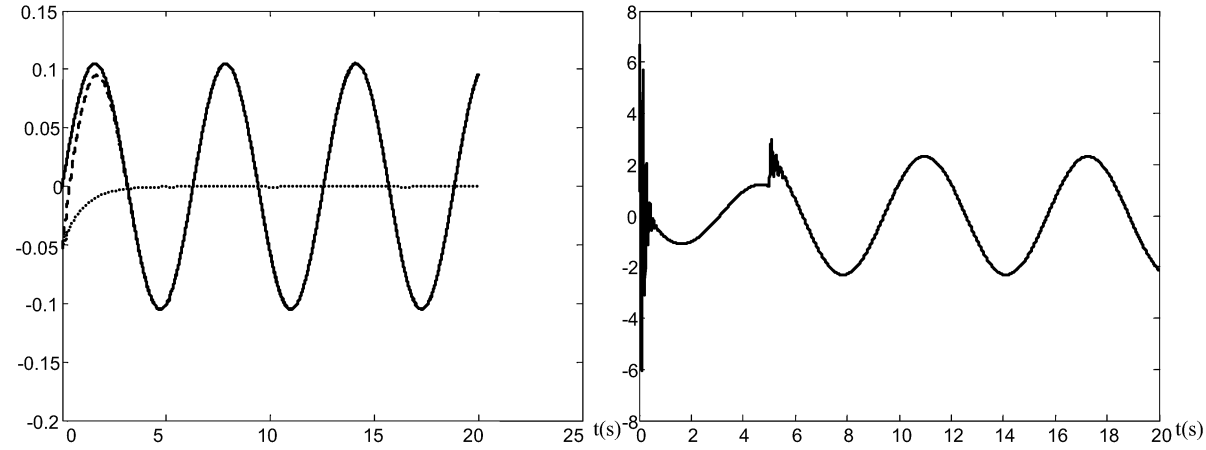


Figure 1: Inverted pendulum system



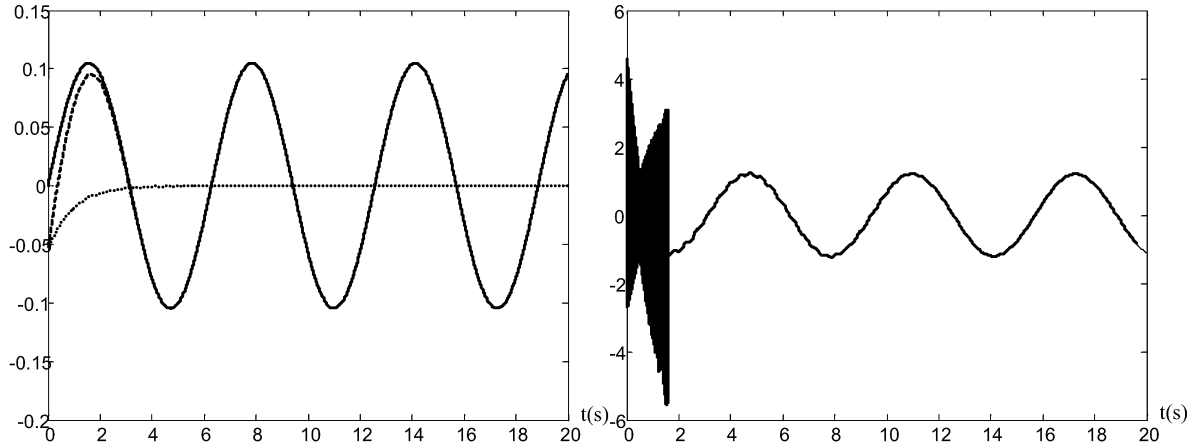
(a) **(b)**

Figure 2: Proposed controller in case of $m_c(t)=1$ **(a) System output** $y(t)$ **(dashed line), its desired value** $y_m(t)$ **(solid line) and** $y(t)-y_m(t)$ **(dotted line) (b) Control input**



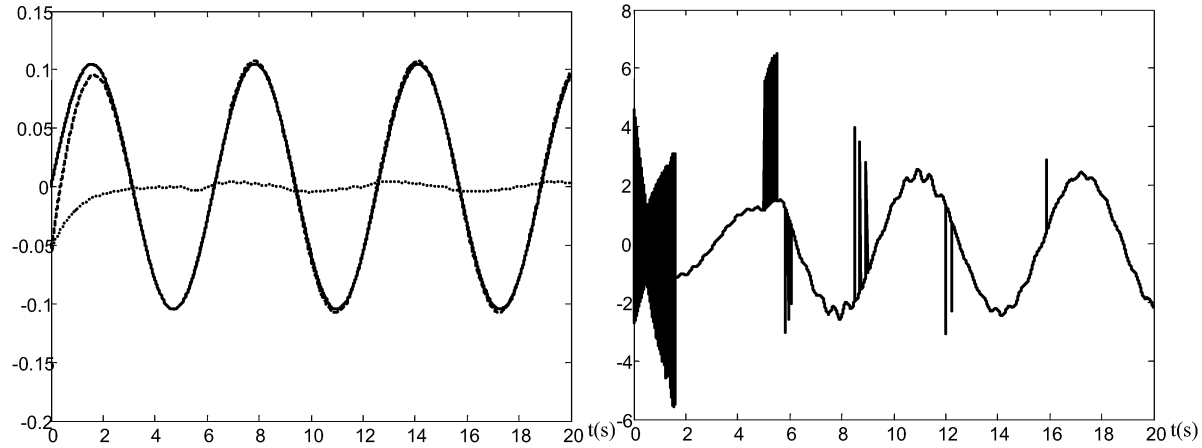
(a) **(b)**

Figure 3: Proposed controller in case of $m_c(t)=\begin{cases}1 \text{ if } t<5s\\ 2 \text{ if } t\geq 5s\end{cases}$ **(a) System output** $y(t)$ **(dashed line), its desired value** $y_m(t)$ **(solid line) and** $y(t)-y_m(t)$ **(dotted line) (b) Control input**



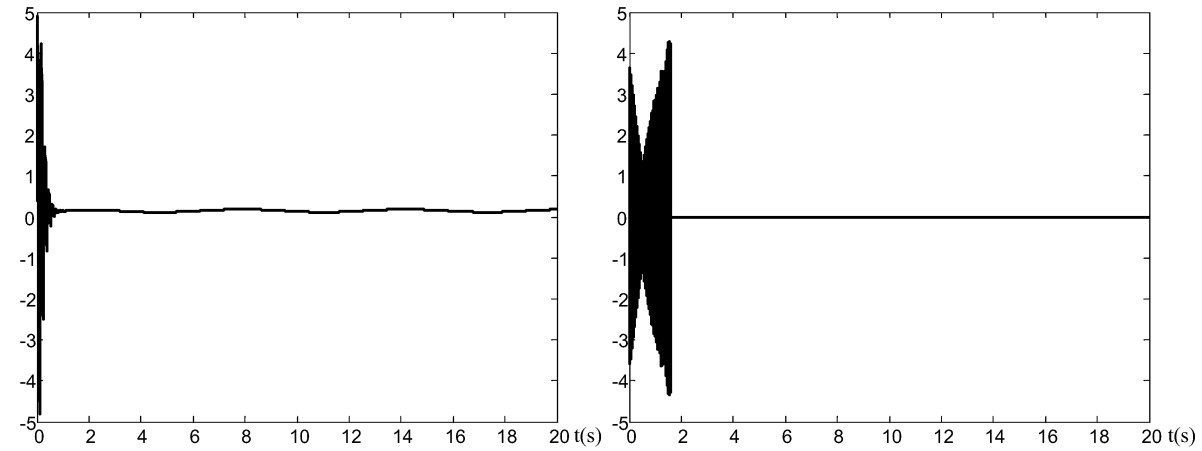
(a) **(b)**

Figure 4: Wang's controller in case of $m_c(t)=1$ (a) System output $y(t)$ (dashed line), its desired value $y_m(t)$ (solid line) and $y(t)-y_m(t)$ (dotted line) (b) Control input



(a) **(b)**

Figure 5: Proposed controller in case of $m_c(t)=\begin{cases}1 \text{ if } t<5s\\2 \text{ if } t\geq 5s\end{cases}$ (a) System output $y(t)$ (dashed line), its desired value $y_m(t)$ (solid line) and $y(t)-y_m(t)$ (dotted line) (b) Control input



(a) **(b)**

Figure 6: Evolution of the robustifying controller in case of $m_c(t)=1$ (a) of the proposed method (b) of the Wang's method

NEURAL NETWORKS IN INDUSTRIAL APPLICATIONS SIMULATION

MODELING OF ABRASIVE FLUID MICRO MACHINING PARAMETERS BY POLYNOMIAL NEURAL NETWORKS AND GENETIC ALGORITHMS

M. Alitavoli
Nariman-zadeh
A. Khalkhali
M. Mehran

June 2005
Engineering Faculty, The University of Guilan
P.O. Box 3756, Rasht, IRAN
E-mail: {tavoli|nnzadeh}@guilan.ac.ir, abolfazl.khalkhali@gmail.com

KEYWORDS: Abrasive Fluid Micro Machining (AFMM), Surface Finish, Material Removal, GMDH, Genetic Algorithms.

ABSTRACT

Micro finishing is considered as an essential step in the improvement of the mechanical properties of metallic components in industry. AFMM is a non-conventional micro machining technique, which is capable of providing excellent surface finish on difficult to reach regions on a wide range of components. In this paper, experimental results of AFMM on aluminium are employed to develop polynomial models using GMDH-type (Group Method of Data Handling) neural networks. Also Genetic Algorithms (GAs) technique are employed for optimal design of connectivity configuration of the GMDH-type neural networks. The aim of such modelling is to show how the material removal and surface finishes varies with the variation of parameters such as number of cycles, and concentration of abrasive. The results of modelling obtained by evolved GMDH-type neural network are then compared with the experimental results. The results have shown a very promising agreement with those obtained from the models.

1. Introduction

In Abrasive Fluid Micro Machining the material is removed by passing a viscous, abrasive fluid media under pressure, through or across a work piece. Generally, the media is extruded through or over the work piece with receprocative. Parameters such as viscosity, pressure, passage size, geometry, and length determine the velocity of the extruded media. Largest grinding forces are observed in the areas that have the greatest restriction. Main abrasive types used in the process are aluminium oxide, boron carbide, silicon carbide and diamonds. The AFMM process acts in a manner similar to grinding or lapping where the extruded abrasive media gently hones edges and surfaces. It is particularly useful when applied to work pieces containing passageways or holes that are considered to be inaccessible with conventional debarring and polishing tools (Rhoades, L. J. 1988, Benedict, G. F. 1987).

The unique features of AFMM such as versatility, efficiency, and economy makes the process useful to perform a wide range of precision micro machining operations in the aerospace and automobile industries, manufacture of dies and medical instruments (Rhoades, L. J. 1988.

Williams and Rajurkar showed that media viscosity and extrusion pressure significantly determine both surface roughness and the material removal rate. They indicated that the major improvement in surface finish takes place within the first few cycles (Williams, R. E. 1989). However, each of these studies considered only a subset of the process parameters and ignored other critical parameters.

In this work, an optimum mathematical equation is found as a model between the material removal, number of cycles and concentration of abrasive as well as between surface finishes, number of cycles and concentration of abrasive. Therefore, it can be seen that the optimum design procedure requires such model identification of AFMM. Theoretically, in order to model a system, it is required to understand the explicit mathematical input-output relationship precisely. Such explicit mathematical modelling is, however, very difficult and is not readily tractable in poorly understood systems. Alternatively, soft-computing methods, which concern computation in imprecise environment, have gained significant attention.

The main components of soft computing, namely, fuzzy-logic, neural network, and genetic algorithm have shown great ability in solving complex non-linear system identification and control problems. Several research efforts

have been expended to use evolutionary methods as effective tools for system identification (Sanchez et al. 1997; Iba et al 1996). Among these methodologies, Group Method of Data Handling (GMDH) algorithm is self-organizing approach by which gradually complicated models are generated based on the evaluation of their performances on a set of multi-input-single-output data pairs (i=1, 2, …, M). The GMDH was firstly developed by Ivakhnenko (Ivakhnenko 1971) as a multivariate analysis method for complex systems modelling and identification. In this way, GMDH was used to circumvent the difficulty of knowing a priori knowledge of mathematical model of the process being considered. In other words, GMDH can be used to model complex systems without having specific knowledge of the systems. The main idea of GMDH is to build an analytical function in a feed-forward network based on a quadratic node transfer function (Farlow 1984) whose coefficients are obtained using regression technique. In fact, real GMDH algorithm in which model coefficients are estimated by means of the least squares method has been classified into complete induction and incomplete induction, which represent the combinatorial (COMBI) and multilayered iterative algorithms (MIA), respectively (Farlow 1984). In recent years, however, the use of such self-organizing network leads to successful application of the GMDH-type algorithm in a broad range area in engineering, science and economics (Farlow 1984; Nariman-zadeh et al. 2003).

In this work, 20 pattern numbers have been obtained from the experiments. Inputs for the network were the number of cycles and concentration of abrasive, while the outputs were material removal and surface finishes. GMDH-type neural networks will then find the input-output relationship in the form of polynomial models for these parameters. Such neural network identification process needs, in turn, some optimisation methods to find the best network architecture. In this way, genetic algorithms are deployed in a new approach to design the whole architecture of the GMDH-type neural networks, i.e., the number of neurons in each hidden layer and their connectivity configuration. The connectivity configuration is not limited to the adjacent layers, as have already proposed in (Nariman-zadeh et al. 2003), so that general structure GMDH-type neural networks, in which neurons in any layers can be connected to each other, are evolved.

2. Experimental Setup

The experiments were performed with alloy aluminium. The samples are shown in Figure 1.

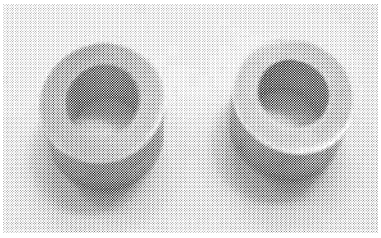


Fig.1 Test specimens of aluminium

The key to successful application of the process lies in the proprietary media formulations. Several years of research, development, testing, and production utilization have lead to unique controllable formulations. Simple laboratory adjustment of formulations can produce a vast range of finishing characteristics. Formulations vary in aggressiveness, radius forming capability, and theology. The three components basic to all media formulations are: (1) Media base material or vehicle, (2) The abrasive grain, and (3) Proprietary additives.

The media used for the present experimentation is a mixture of putty, abrasive particles (Sic), and hydraulic oil. Figure 2 shows the whole setup of AFMM machine. The basic elements of the machine tools are as follows: **1**.Hydraulic Power Unit **2**.Media Chamber Assemblies **3**.Fixture **4**.Support Structure **5**.Control System.

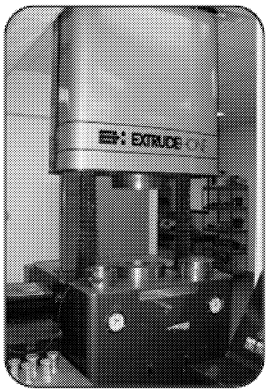


Fig.2 Abrasive Fluid Micro Machine

The mechanism of AFMM process is shown in Figure 3. A hydraulic unit develops the fluid pressure required by the hydraulic clamping cylinders and media extrusion rams. Standard units operate within the 0-1600 p.s.i. Current machines are designed with two vertically opposed media chamber assemblies. The assemblies consist of a honed cylinder, a specially designed piston and cup seal subassembly, and a hydraulic ram. Diameters and lengths of the honed cylinders determine the volumetric capacity of a given machine. Hydraulic clamping cylinders are mounted in the base and attached to head structure.

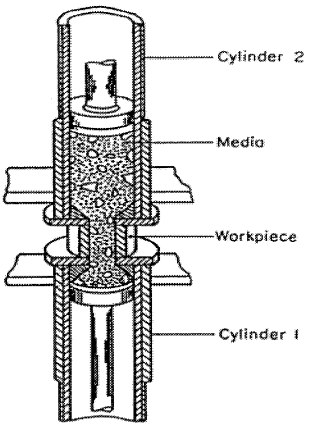


Fig.3 Schematic AFMM Process

3. Modeling Using GMDH Neural Networks

By means of GMDH algorithm a model can be represented as set of neurons in which different pairs of them in each layer are connected through a quadratic polynomial and thus produce new neurons in the next layer. Such representation can be used in modelling to map inputs to outputs. The formal definition of the identification problem is to find a function $\hat{f}$ so that can be approximately used instead of actual one, f, in order to predict output $\hat{y}$ for a given input vector $X = (x_1, x_2, x_3, ..., x_n)$ as close as possible to its actual output y. Therefore, given M observation of multi-input-single-output data pairs so that

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \quad (i=1,2,...,M), \tag{1}$$

it is now possible to train a GMDH-type neural network to predict the output values $\hat{y}_i$ for any given input vector

$$X = (x_{i1}, x_{i2}, x_{i3}, ..., x_{in}), \text{ that is}$$
$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \qquad (i=1,2,...,M). \tag{2}$$

The problem is now to determine a GMDH-type neural network so that the square of difference between the actual output and the predicted one is minimised, that is

$$\sum_{i=1}^{M} [\hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) - y_i]^2 \to \min . \tag{3}$$

General connection between inputs and output variables can be expressed by a complicated discrete form of the Volterra functional series in the form of

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + ... \tag{4}$$

which is known as the Kolmogorov-Gabor polynomial (Sanchez et al. 1997; Iba et al. 1996; Ivakhnenko 1971;

Farlow 1984; Nariman-zadeh et al. 2003). This full form of mathematical description can be represented by a system of partial quadratic polynomials consisting of only two variables (neurons) in the form of

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{5}$$

In this way, such partial quadratic description is recursively used in a network of connected neurons to build the general mathematical relation of inputs and output variables given in equation (4). The coefficients a_i in equation (5) are calculated using regression techniques (Farlow 1984; Nariman-zadeh et al. 2003) so that the difference between actual output, y, and the calculated one, $\hat{y}$ for each pair of x_i, x_j as input variables is minimized. Indeed, it can be seen that a tree of polynomials is constructed using the quadratic form given in equation (5) whose coefficients are obtained in a least-squares sense. In this way, the coefficients of each quadratic function G_i are obtained to

Optimally fit the output in the whole set of input-output data pair, that is

$$E = \frac{\sum_{i=1}^{M} (y_i - G_i())^2}{M} \to \min \tag{6}$$

In the basic form of the GMDH algorithm, all the possibilities of two independent variables out of total n input variables are taken in order to construct the regression polynomial in the form of equation (5) that best fits the dependent observations (y_i, $i=1, 2, ..., M$) in a least-squares sense. Consequently, $\binom{n}{2} = \frac{n(n-1)}{2}$ neurons will be built up in the first hidden layer of the feedforward network from the observations $\{(y_i, x_{ip}, x_{iq})$; $(i=1, 2, ..., M)\}$ for different $p, q \in \{1,2,...,n\}$. In other words, it is now possible to construct M data triples $\{(y_i, x_{ip}, x_{iq})$; $(i=1, 2, ..., M)\}$ from observation using such $p, q \in \{1,2,...,n\}$ in the form

$$\left[\begin{array}{cc|c} x_{1p} & x_{1q} & y_1 \\ x_{2p} & x_{2q} & y_2 \\ \hline x_{Mp} & x_{Mq} & y_M \end{array}\right].$$

Using the quadratic sub-expression in the form of equation (5) for each row of M data triples, the following matrix equation can be readily obtained as

$$A\,\mathbf{a} = Y \tag{7}$$

where $\mathbf{a}$ is the vector of unknown coefficients of the quadratic polynomial in equation (5)

$$\mathbf{a} = \{a_0, a_1, a_2, a_3, a_4, a_5\} \tag{8}$$

and

$$Y = \{y_1, y_2, y_3, ..., y_M\}^T \tag{9}$$

is the vector of output's value from observation. It can be readily seen that

$$A = \left[\begin{array}{cccccc} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \hline 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{array}\right] \tag{10}$$

The least-squares technique from multiple-regression analysis leads to the solution of the normal equations in the form of

$$\mathbf{a} = (A^T A)^{-1} A^T Y \tag{11}$$

which determines the vector of the best coefficients of the quadratic equation (5) for the whole set of M data triples. It should be noted that this procedure is repeated for each

neuron of the next hidden layer according to the connectivity topology of the network. However, such a solution directly from normal equations is rather susceptible to round off errors and, more importantly, to the singularity of these equations.

4. Application of GA in the Topology Design of GMDH-Type Neural Networks

Such stochastic methods are commonly used in the training of neural networks in terms of associated weights or coefficients that have successfully performed better than traditional gradient-based techniques (Nariman-zadeh et al. 2003). The literature shows that a wide range of evolutionary design approaches either for architectures or for connection weights separately, in addition to efforts for them simultaneously (Yao 1999) In the most GMDH-type neural network, neurons in each layer are only connected to neuron in its adjacent layer as it was the case in Methods I and II previously reported in (Nariman-zadeh 2003). Taking this advantage, it was possible to present a simple encoding scheme for the genotype of each individual in the population as already proposed by authors (Nariman-zadeh 2003). The encoding schemes in generalized GMDH neural networks must demonstrate the ability of representing different length and size of such neural networks.

In Figure 4, neuron *ad* in the first hidden layer is connected to the output layer by directly going through the second hidden layer. Therefore, it is now very easy to notice that the name of output neuron (network's output) includes *ad* twice as *abbcadad*. In other words, a virtual neuron named *adad* has been constructed in the second hidden layer and used with *abbc* in the same layer to make the output neuron *abbcadad* as shown in the Figure 4. It should be noted that such repetition occurs whenever a neuron passes some adjacent hidden layers and connects to another neuron in the next 2^{nd}, or 3^{rd},or 4^{th},or … following hidden layer. In this encoding scheme, the number of repetition of that neuron depends on the number of passed hidden layers, ñ, and is calculated as $2^{ñ}$. It is easy to realize that a chromosome such as *abab bcbc*, unlike chromosome *abab acbc* for example, is not a valid one in GS-GMDH networks and has to be simply re-written as *abbc*.

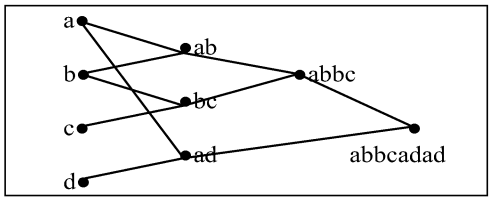


Fig.2 a Generalized GMDH Network Structure of a Chromosome

5. THE AFMM GMDH NEURAL NETWORK MODELLING

The parameters of interest in this multi-input single-output system that affect the material removal and surface finishes are the number of cycles and concentration of abrasive. There has been a total number of 20 input-output experimental data considering 2 input parameters, namely, number of cycles and concentration of abrasive that were obtained from experiments carried out for abrasive fluid micro Machining.

The GMDH-type neural networks are now used for such input-output data to find the polynomial model of material removal and surface finishes in respect to their effective input parameter. In order to genetically design such GMDH-type neural network described in previous section a population of 15 individuals with a crossover probability of 0.7 and mutation probability of 0.07 has been used in 200 generation that no further improvement has been achieved for such population size. The structure of the evolved 2-hidden layer GMDH-type neural network is shown in Figure5, Figure 6 corresponding to the genomes representation of ***aaababab*** for material removal and ***aaabbbbb*** for surface finishes which ***a, b*** stand for number of cycle and concentration of abrasive, respectively.

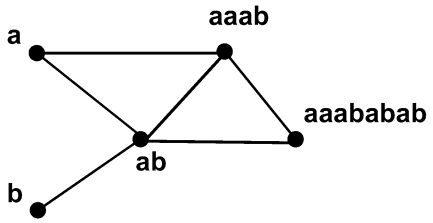


Fig.5 Evolved Structure of Generalized GMDH Neural Network for material removal

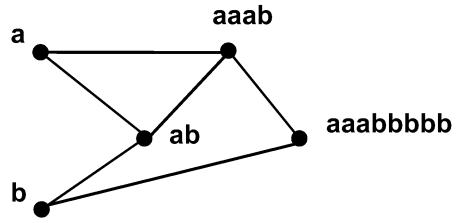


Fig.6 Evolved Structure of Generalized GMDH Neural Network for surface finishes

However, in order to demonstrate the prediction ability of evolved GMDH-type neural networks, the data has been divided into two different sets, namely, training and testing sets. The training set, which consists of 15 out of 20 inputs-output data pairs, is used for training the neural network models the evolutionary method of this paper. The testing set, which consists of 5 unforeseen inputs-output data samples during the training process, is merely used for testing to show the prediction ability of such evolved GMDH-type neural network models during the training process. The structures of GMDH network are same of Figure 5, Figure 6. The very good behaviour of such GMDH-type neural network models are also depicted in Figure 7 and Figure 8. The corresponding polynomial representation of such model for material removal is:

$$Y_1 = -94.832 + 0.0356N + 2.6189C_{ab} - 0.0020N^2 - 0.0178C_{ab}^2 + 0.0079N\,C_{ab} \quad (12)$$

$$Y_2 = 0.318 - 0.9388Y_1 + 0.0051N + 0.0059Y_1^2 + 0.0003N^2 - 0.0030Y_1N \quad (13)$$

$$\text{material removal} = 0.102 + -12.359Y_1 + 13.224Y_2 - 3.945Y_1^2 - 4.486Y_2^2 + 8.4384Y_1Y_2 \quad (14)$$

And surface finishes is:

$$Y_3 = 3.0226 - 0.0076N - 0.0251C_{ab} + 0.0001N^2 + 0.0001C_{ab}^2 + -0.0001NC_{ab} \quad (15)$$

$$Y_4 = -1.119 + 0.0073N + 2.4602Y_3 - 0.465Y_3^2 - 0.0051Y_3N \quad (16)$$

$$\text{surface finish} = -1.385 + 0.0113Y_4 + 2.4238C_{ab} - 0.2394C_{ab}^2 - 0.011Y_4C_{ab} \quad (17)$$

Where N, C_{ab} stand for number of cycle and concentration of abrasive, respectively.

It is clearly evident that the evolved GMDH-type neural network in terms of simple polynomial equations can successfully model and predict the output of testing data that has not been used during the training process.

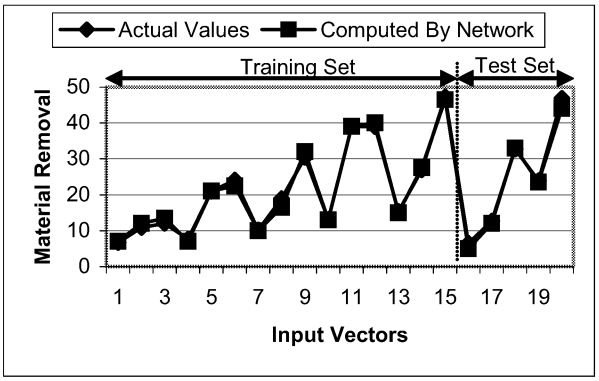


Fig.7 Comparison of experimental material removal with Evolved GMDH Neural (Modelling & Prediction)

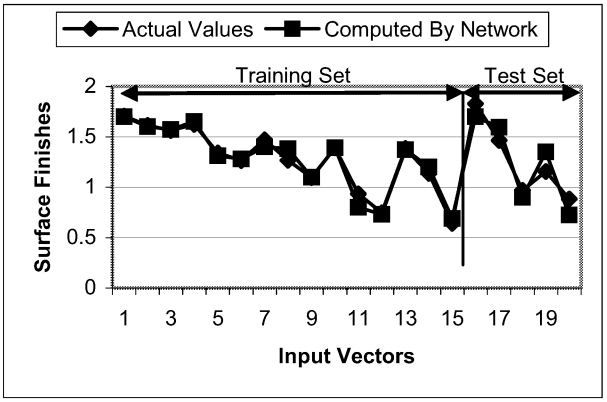


Fig.8 Comparison of surface finishes and Evolved GMDH Neural (Modelling & Prediction)

6. CONCLUSION

Evolutionary methods for designing generalized GMDH-type networks have been proposed and successfully applied to in modelling of AFMM experimental results. In this way, it has been shown that GMDH-type networks provide effective polynomial to model and predict the material removal and surface finishes of the aluminium in AFMM.

Besides, a new encoding scheme has been also presented to genetically design generalized GMDH-type neural networks in which the connectivity configuration in such networks is not limited to adjacent layers, unlike the conventional GMDH-type neural networks.

References:

Benedict, G.F.1987."*Non-Traditional Manufacturing Processes"*, New York, NY: Marcel Dekker Inc. pp. 53-65.

Farlow, S.J., ed. 1984. *"Self-organizing Method in Modelling: GMDH type algorithm"*, *Marcel Dekker Inc*.

Iba, H.; H. deGaris; and T. Sato. 1996.A "numerical Approach to Genetic Programming for System Identification", *Evolutionary Computation* 3(4): 417-452.

Ivakhnenko, A.G. 1971. "Polynomial Theory of Complex Systems", *IEEE Trans. Syst. Man & Cybern*, SMC-1, 364-378.

Nariman-Zadeh, N.; A. Darvizeh; and R. Ahmad-Zadeh; 2003."*Hybrid Genetic Design of GMDH-Type Neural Networks Using Singular Value Decomposition for Modelling and Prediction of the Explosive Cutting Process*", *Proceedings of the I MECH E Part B Journal of Engineering Manufacture*, Vol. 217, pp. 779 – 790.

Rhoades, L. J.1988, "Abrasive Fluid Micro Machining," *Manufacturing Engineering*, Vol. 101, pp. 75-78.

Sanchez E.; T. Shibata; and L. A. Zadeh. 1997. "*Genetic Algorithms and Fuzzy Logic Systems*", World Scientific.

Williams, R. E. and Rajurkar, K. P. 1989, "Metal Removal and Surface Finish Characteristics in Abrasive Fluid Micro Machining," *Mechanics of Deburring and Surface Finishing Processes* (R. J. Stango and P.R. Fitzpatrick, eds.), New York, NY: ASME, Vol. 38. pp. 93-106.

Yao, X. 1999. "*Evolving Artificial Neural Networks*", *Proceedings of IEEE*, 87(9):1423-1447.

Characterization of petroleum wells using fuzzy systems

Rafika HAMZA, Khaled MECHICHI, Chokri REKIK, Abdellaziz ABOUD and Nabil DERBEL
Research unit on Intelligent Control, design and Optimisation of complex Systems (ICOS)
University of Sfax, ENIS, BP. W, 3038, Sfax, Tunisia.
Phone: (216-4) 274.088, Fax. (216-4) 275.595.
Emails: nabil.derbel@ieee.org

Abstract: The object of this paper concerns the determination of the percentage of the rocks in oil wells. The object is to estimate the existence of oil pockets.
In fact, the current method of characterization of an oil tank is based on human experiments and on empirical relations which can generate errors. So, in this paper, we propose an approach which can limit these errors by using fuzzy systems as universal approximators. This approach can give an idea on the existence of different type of rocks by specifying the percentage of each one from them. These results facilitate the search of the zones which can contain oil.

Keywords: fuzzy systems, petroleum wells, lithology, rocks.

1. Introduction:

The data analysis plays an important role on new engineering sciences. The data analysis approach used should be able to provide a reasonable summary as well as data analysis. There are two main of categories of data analysis: descriptive and inferential (Phylips and Sincich 1998).
- Descriptive data analysis is simply aims to find a description of the data as presented solution. No prediction of what might have been achieved outside the range is expected, nor should it be undertaken.
- Inferential analysis, however, the analysis tool is expected to derive the underlying function from which the data derive and therefore allows the prediction of data that could be expected in the experiment for different input values [7, 8, 9].

For more reliable data analysis; new approaches will be adopted. It is required to offer reasonable interpolation performance and provide some indication when extrapolation is appropriate.
In petroleum industry well log data analysis among to class of non-linear data analysis techniques (Mendenhall and Sincish 1992). Despite of innovative probabilistic methods for dealing with some limited cases of nonlinear data; it is no yet able to analyse complicated data especially where the nature of the non-linearity is unknown which becomes an identification problem of “black box” (model) which associates known input to their measured corresponding output then problem is to determine a function that describes the relationship between the input and the output.

The cost of developing a petroleum tank studies becomes very expensive especially for marginal field or deepwater tank. The first step starts by drilling exploration wells at different locations around the region believed to hold the tank. Then well logging instruments are then lowered into each borehole to collect data at different depth. These data are called well log data, which will be intensively evaluated to predict tank's potential and parameter. Well logging fall into three categories: electrical, nuclear and acoustic and there are over fifty different types of logging measure in addition to direct borehole rock sampling for further laboratory analysis called “core data” which it is the most accurate way of well hydrocarbon potential. However core data are very difficult and expensive to obtain [1, 2, 3, 4].

The uses of fuzzy logic and more recently hybrid GA fuzzy Logic (FL) simplifies the development of an intelligent data analysis tool with the following features:
- Human rich experience could be incorporated into the fuzzy knowledge base (rules base) which is very close to human natural language. - The model input data or knowledge base is not necessarily precise and complete.
- Partially matched conclusions could be inferred from the fuzzy facts and the established fuzzy knowledge base.

This approach is suitable to the application of most large data analysis problems as the model for each case may vary greatly and it allows the incorporation of intelligent and human knowledge

to deal with each specific case. However, the extraction of fuzzy rules on the data analysis process could be a very difficult task to human analysts which could be considered as the major disadvantage of fuzzy logic in large data analysis especially for well log data analysis.

1. Problems formulation:

The current method for the identification of the lithology is based on relations as written below [5, 6]:

- **Dolomite** weak gamma ray + sonic weak.
- **Evaporate** weak gamma ray + sonic strong with strong resistivity.
- **Limestone** very weak gamma ray + sonic very weak with average resistivity
- **Shale** gamma ray more than the average + sonic very strong

Using these relations, the geologist tries to identify the lithology of the well basing on his knowledge. After that He use empirical relations to characterize the well by giving:

- its volume
- its porosity
- its rate of saturation
- its permeability
- the various fluids which it contains (oil, gas, water).

To illustrate this method, we give the example of figure 1.

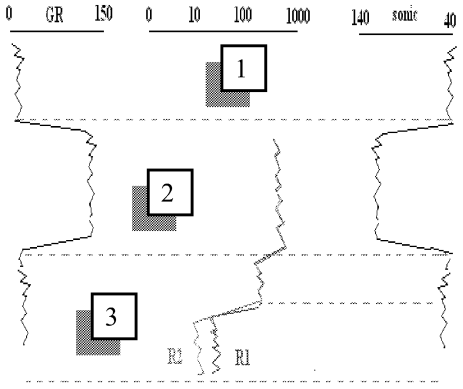


Fig.1 Diagram showing the current method

In zone 1, the value of measurement gamma ray is weak; the same thing for sonic measurement, then the layer is composite by the Dolomite rock on this depth. In the same way, for zone 2, the value of measurement gamma ray is more than the average and for a very strong value of sonic measurement, so, the rock is Shale. In zone 3, the measurement of gamma ray is very weak; the sonic measurement is very weak also with both average resistivities, so the rock is Limestone. These results are sullied by errors due to the several factures: the human expert, reading, and interpretation.

Moreover, results of such analysis, mud log or core data, are given in a form which does not reflect in any case the reality: in a given depth the layer is formed by only one rock.

The current method of the characterization of the oil well is not very exact and does not give acceptable results. In what follows, we will propose an approach which will limit these errors by using the fuzzy systems as universal approximateurs. This approach makes possible to give an idea on the existence of various types of rocks by specifying a percentage of each one of them. One takes as example of studies the wells of the field of Guebiba.

2. Development of the fuzzy model

Each value of a measurement of the logging is limited [10, 11].

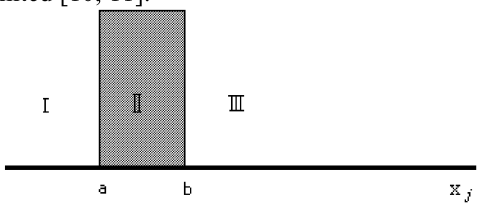


Fig.2. Degree of existence of a rock according to a value

According to this value, we can deduce the existence or not of a rock: if x_j belongs to interval II then the rock exists, elsewhere it does not exist. This traditional interval will be transformed into an interval which brings closer the result with reality.

2.1. Function of membership :

We choose the Gaussian function as being related to membership to simulate the fuzzy model. The expression of the Gaussian function is:

$$G_{m,\sigma} = \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

m : the center of function Gaussian

σ : Opening of the Gaussian function which will be defined by the following equations for an attenuation of 1/n of the function $G_{m,\sigma}$.

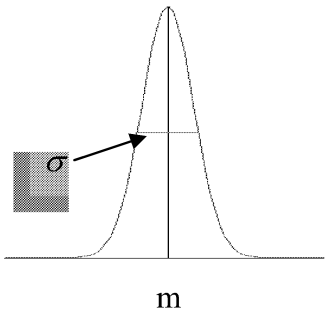


fig.3. The Gaussian function

To seek the opening of Gaussian, we will fix the dispatcher L for an attenuation of 1/n, (n >1).
If we resolve this equation:

$$G_{m,\sigma} = \frac{1}{n}$$

We obtain:

$$\sigma = \frac{l}{2\sqrt{2\log(n)}}$$

If we choose (x – m) = l = b – a

And n = e^1, the expression of σ becomes:

$$\sigma = \frac{l}{2\sqrt{2}}$$

Each traditional interval will be transformed into fuzzy interval. We obtain intervals of the following form:

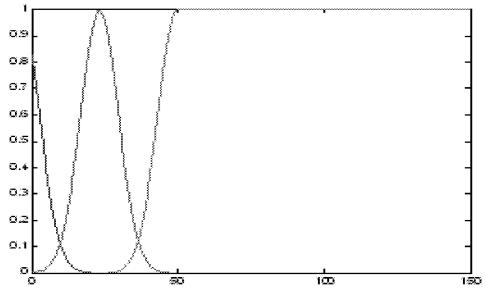


fig.4. Fuzzy interval represented by the gaussian function

3.2. Parameters of entry:

The parameters of entry are three, which represent measurements of logging. This choice is not random; these three parameters are the most common in the databases available.
The three tests are: the gamma ray (GR.), sonic (DT) and density (RHOB).

3.3. Parameters of exit:

A parameter of exit is defined as being the percentage of existence of rocks i (i=1:14) according to the depth.
In this work, we are interested only in six great families of rock subdivided on 14 under families.

Table.1. Families and under families of rocks considered

Families	Under families
SHALE	MARINE
	NON –MARINE
COAL	INFERIOR
	LIGNITE
	ANTHRACITE
	BITUMINOUS
SANDSTONE	POROUS
	TIGHT
SILISTONE	
EVAPORITES	SALT
	GYPSUM
	ANHYDRITE
LIMESTONE	POROUS
	TIGHT

Based on the model of Takagi-Sugeno (TSK) of order 0, the output of the fuzzy model is the balanced sum of the exits of each rule with degree of applicability of the premise of each rule of the form: If-Then

R_{ljk} : if (GR is A_l) and (DT is B_j) and (RHOB is C_k) then $z = y_{ljk} = y_l \times y_j \times y_k$

Avec : R_l = if (GR is A_l) then $z = y_l$

R_j =if (DT is B_j) then $z = y_j$

R_k = if (RHOB is C_k) then $z = y_k$

During treatment, we calls the method of inference max-prod where the operator ' AND ' is represented by the function ' min ' and the implication introduced by ' Then ' is represented by the function ' produced ' The exit is presented then by the following expression:

$$z_i = \frac{\sum_l \sum_j \sum_k \mu_l \eta_j \alpha_k y_{ljk}}{\sum_n \sum_v \sum_s \mu_n \eta_v \alpha_s}$$

μ_l : weight of the rule R_l ; $\mu_l = \mu_{A_l}$ (GR).

η_j : weight of the rule R_j ; $\mu j = \mu_{B_j}$ (DT).

α_k : weight of the rule ; $\mu_k = \mu_{C_k}$ (RHOB).

4. Simulation:

The simulation part will relate two problems of the identification of the lithology of the oil wells:

- The first will be the identification of the composition of the rocks according to the depth.
- The second will be the research of the zones which can be oil pockets.

The data bases, provided by the TPS, represent the answers to the tests of logging applied to the various wells.
Moreover documentation concerning the intervals of answers of the rocks at the various tests is incomplete, it does not cover all the families of existing rocks in Tunisia.
Therefore the provided data bases, do not cover the entire well, in certain cases they represent only a negligible part of the well. These problems make the development of a general model difficult.

4.1. Identification of the composition of the rocks

One will treat the fuzzy model corresponding to various entries in order to note the influence of the increase in the number of entry of the model on the precision of the results.
We take the case of well GUEBIBA 1, we present at the entry of the model combination of the following tests: GR, DT and RHOB.

4.1.1. Case of a system at only one entry

Si (GR est A_i) Alors ($z = y_i$)

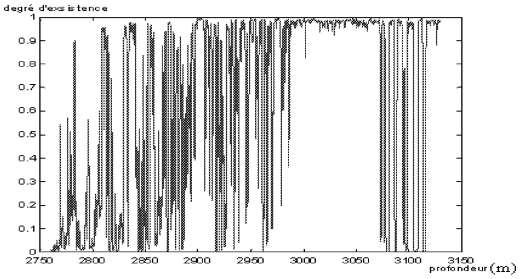


fig.5. Degree of existence of the rock limestone porous according to GR (before filtering).

Fig.5. represents the existence of the rock limestone porous in the tank part of GB1 well. It is illegible because of the existence of limit of measurement. We use a filter of the type passes low from order 1 whose equation is:

$$H(z) = \frac{(1-a)\times z^{-1}}{1-a\times z^{-1}} \qquad \text{with : a=0.99}$$

We applied this filter to the output of our model.

The following curve is obtained:

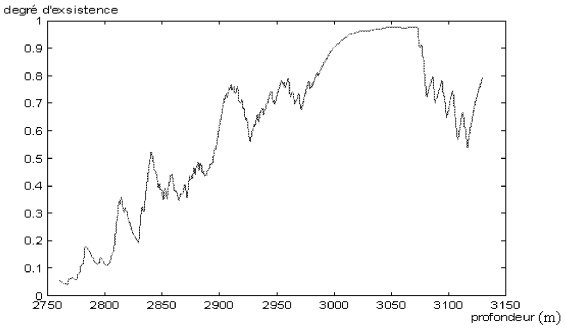


fig.6. Degree of existence of the rock limestone porous according to GR (after filtering).

The figure obtained became more readable and likely to be used for the identification of degree of existence of this rock in this well.

4.1.2. Case of a system at 2 entries

if (GR is A_l) and (DT is B_j) then ($z = y_{lj}$)

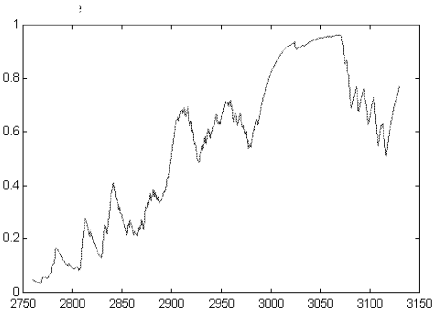


fig.7. Degree of existence of the rock limestone porous according to GR and DT (after filtering).

4.1.3. Case of system at 3 entries

if (GR is A_l) and (DT is B_j) and (RHOB is C_k)

then ($z = y_{ljk}$)

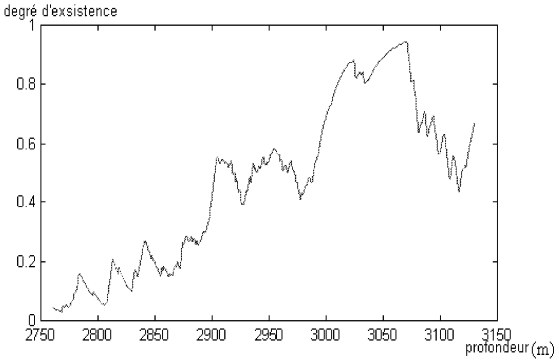


fig.8. Degree of existence of the rock limestone porous according to GR, RHOB and DT (after filtering).

Interpretation

- By increasing the number of entry of our fuzzy model, the precision increases. The curve translating the lithology of the well becomes increasingly close to reality.
- This approach allows the identification of the degree of existence of the rocks according to the depth.

4.2. The research of the oil pockets

In this part, we are interested in another problem which relates to oil industry: it is the localization of the pockets which contain oil. The model seeks structures having the form of the figure 9.

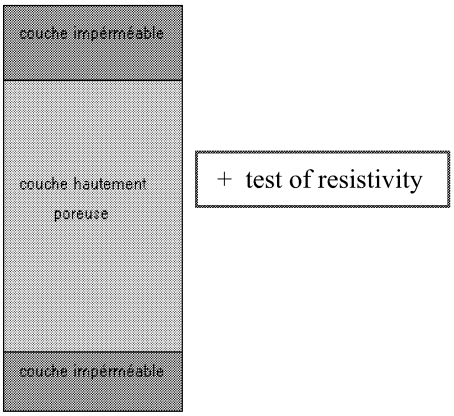


fig.9. Structure of oil pocket adapted to model

5. Conclusion

These conclusions are at the same time the strong points which characterize our approach compared to the current method used in the characterization of lithology in the oil field.

Like first point, the passage from traditional interval to fuzzy interval gave the exit in the form of percentage of rock contrary to the current method which supposes that with a given depth the ground is formed by only one rock.

By using this approach, the identification of lithology became more precise by correcting the error due to the personal element. Also, the manipulation of the parameters of logging obtained at the time and after drilling became easier and more rapid, not requiring a personal element highly qualified. Moreover the model is able to identify the parts of the well whose configuration corresponds to that of a tank without specifying if there is oil or not, for lack of lack of data during the development of this model.

If one is interested in the model, each time that one increases the number of entry the curves obtained becomes increasingly representative of reality. Each new entry thus corresponds to a new test of diagraphics for each new entry one will take into account new properties of the rocks. The model becomes more sensitive then and gives proportions of more reasonable rocks. But each time that one increases the number of entries the base of rules increases.

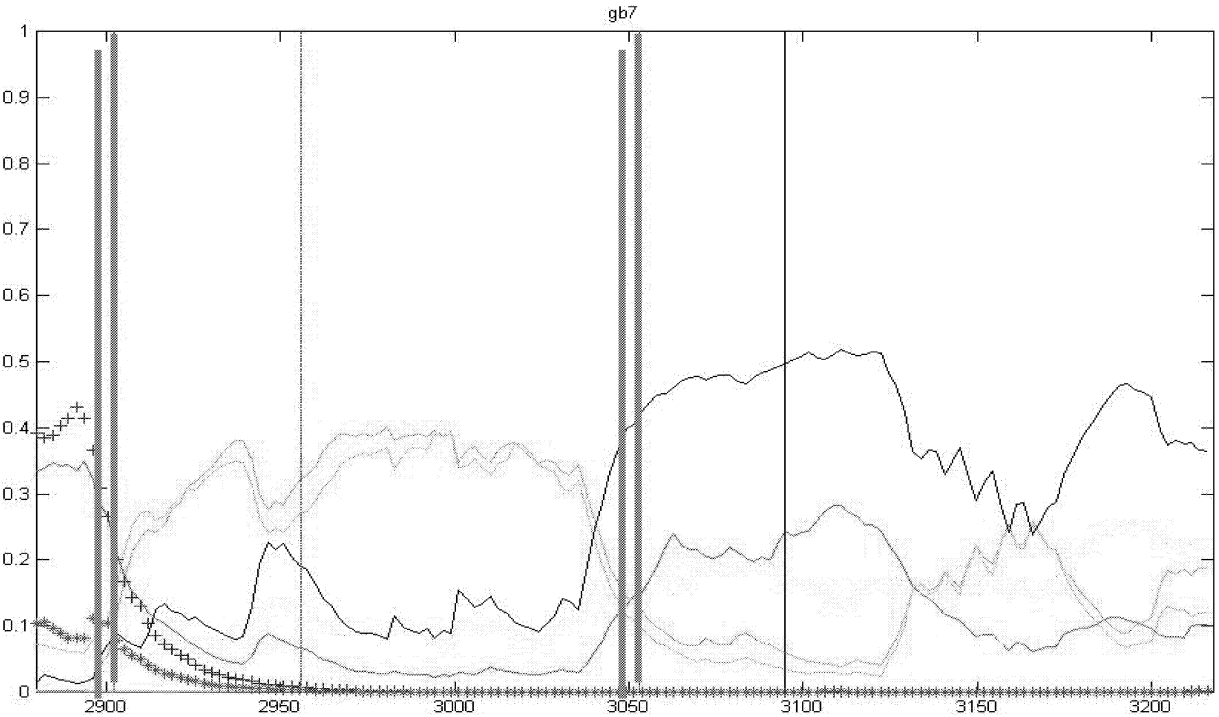


Fig.10: Identification of a pocket in a well of the fields of Guebiba

[1] Ashok Kumar Belani. August 1988. ESTIMATION OF MATRIX BLOCK SIZE DISTRIBUTION IN NATURALLY FRACTURED RESERVOIRS.

[2] B. G. Kelkar, R. F. Gamble, D. R. Kerr,L. G. Thompson,S. Shenoi. August 31, 1997. APPLICATION OF ARTIFICIAL INTELLIGENCE TO RESERVOIR CHARACTERIZATION: AN INTERDISCIPLINARY APPROACH.

[3] B., Mohaghegh, S., Ameri, September, 1995. SPE 30978. West Virginia University. State-Of-The-Art in Permeability Determination From Well Log Data.

[4] Benjamin Chin Ming Fung. Simon Fraser University, 1999. HIERARCHICAL DOCUMENT CLUSTERING USING FREQUENT ITEMSETS.

[5] Bension Sh. Singer and A. Green. Australian Mineral Exploration, Sydney, Australia. Transient field solution for a layer of finite thickness on a resistive basement.

[6] Brainard and Robert Glass. University of Wisconsin, Madison. A HYBRID HYDROLOGIC-GEOPHYSICAL INVERSE TECHNIQUE FOR THE ASSESSMENT AND MONITORING OF LEACHATES IN THE VADOSE ZONE.

[7] C.D. Morgan, K. P. McClure, S. Robert Bereskin, Milind D. Deo, Kevin Weller. RESERVOIR CHARACTERIZATION OF THE LOWER GREEN RIVER FORMATION, SOUTHWEST UINTA BASIN, UTAH.

[8] C.E. Romero, J.N. Carter, R.W. Zimmerman, and A.C. Gringarten. SPE 62942. Imperial College of London UK.

[9] Christopher Skelt. UNOCAL Corporation, Houston, USA. PREDICTING ACOUSTIC IMPEDANCE IN SAND-CLAY-SILT SYSTEMS.

[10] D.G. Chen, N.B. Hargreaves, D.M. Ware, and Y. Liu. June 30, 2000. University of Washington. A fuzzy logic model with genetic algorithm for analyzing fish stock–recruitment relationships.

[11] J. Yen, R. Langari and L. Zadeh: "Industrial Applications of Fuzzy Logic and Intelligent Systems". New York, IEEE Press, 1995.

Optimal Control of a Container Crane by A Fuzzy Logic Controller, Robustness and Stability

Najla KRICHEN MASMOUDI, Mohamed DJEMEL and Chokri REKIK

Research unit *on Intelligent Control, design and Optimisation of complex Systems (ICOS)*
University of Sfax, ENIS, BP. W, 3038, Sfax, Tunisia.
Phone: (216-74) 274.088, Fax. (216-4) 275.595.
Emails: najla_krichen@yahoo.fr, mohamed.Djemel@enis.rnu.tn , chokri.rekik@enis.rnu.tn

Abstract: The objective of this study concerns the determination of the optimal control strategies of a rotary crane. Such system is represented by a sixth order nonlinear mathematical model. The proposed approach consists of the decomposition of the system into two interconnected subsystems which are linear with respect to their state and control vectors. The non linearities are located in the interconnexion terms. In this context, local feedback gains can be computed by solving a nonlinear Riccati equation, whereas, the local gains depend on the state vector of the other subsystem. Within this approach, a Takagi Sugeno fuzzy logic system has been constructed in order to identify these gains. Simulation results show that the proposed approach yields to satisfactory performances.

Keywords: Rotary crane; Optimal control; Differential Dynamic Programming; Fuzzy logic control; Stability.

1- INTRODUCTION:

The interest of this work is focused upon the determination of optimal control laws for nonlinear systems.

Usually, the used method is the classical differential dynamic programming procedure (DDP). It can be described as a hybrid decomposition procedure (recurrent and iterative) for resolving problems of nonlinear systems. However, the major drawback of this method is that the optimal solution depends on the choice of the initial trajectory. In fact, if this trajectory is far from the optimal one, this leads to the convergence to local minimums. In some cases, it can diverge.

This paper is divided into five sections. The second section presents the problem formulation. In the third section, the decomposition of the system is presented. The fourth section is focused upon the proposed approach that satisfy the desired performances. Finally, the last section study the robustness and the stability of the system.

2- Problem formulation:

The fundamental motion of a rotary crane is the rotation and load hoisting [9,10]. For simplicity, we assume that the container load can be regarded as a material point and that frictional torques which may exist in torque-transfer mechanisms can be neglected.

The following notations, which are shown in figure 1, will be used: θ_1, angle of rotation of the trolley drive motor; J_1, total moment of inertia of trolley drive motor, a brake, a drum and a set of reduction gears; b_1, equivalent radius of the drum of trolley drive motor which is reduced to the motor side; θ_2, angle of rotation of the hoist motor; J_2, total moment of inertia of hoist motor, a brake, a drum and a set of reduction gears; b_2, equivalent radius of the drum of the hoist motor which is reduced to the motor side; ϕ, the load swing angle; m, total mass of the trolley and operator's cab; M, total mass of the contain load, the spreader, and the attached equipment; T_1, driving torque generated by the trolley drive motor; T_2, driving torque generated by the hoist motor; and g, the acceleration of gravity.

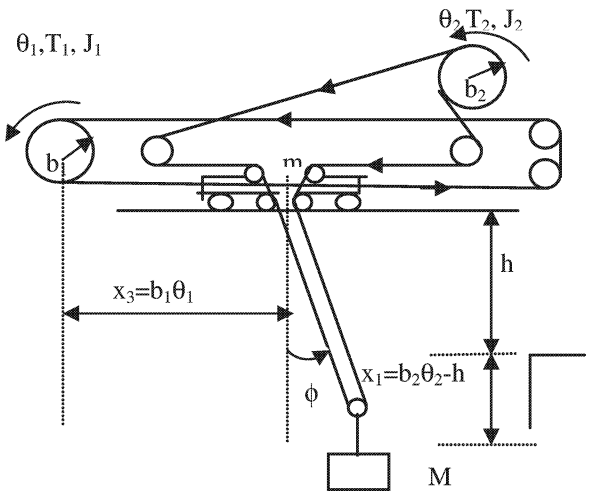


Figure 1: Schematization for the rotary crane

The state variables are x_1, x_2, x_3 by: $x_1=b_2\theta_2$, $x_3=b_1\theta_1$ and $x_5=\phi$

The discretization of the system was done using Euler approach with a sample time T=0.2s.

The process to be controlled can be described by a sixth-order nonlinear differential system, described by

$$\begin{cases} x_{1,k+1} = x_{1,k} + T\, x_{2,k} \\ x_{2,k+1} = x_{2,k} + T\, u_{1,k} \\ x_{3,k+1} = x_{3,k} + T\, x_{4,k} \\ x_{4,k+1} = x_{4,k} + T\,(u_{2,k} + \delta_1\, x_{5,k}) \\ x_{5,k+1} = x_{5,k} + T\, x_{6,k} \\ x_{6,k+1} = x_{6,k} - \dfrac{T}{(x_{1,k}+h)}(u_{2,k} + \delta_2\, x_{5,k} + 2\, x_{2,k}\, x_{6,k}) \end{cases} \quad (1)$$

Parameters δ_1 and δ_2 are given by $\delta_1=67$ and $\delta_2=77$.

The objective is to determine the optimal strategies of such system, in order to transfer the rotary crane from any initial state to the desired state described by: $X_d=[0\ 0\ 0\ 0\ 0\ 0]^T$.

We choose the following quadratic criterion defined as:

$$J = \frac{1}{2}\sum_{i=1}^{200}(X^T Q X + U^T R U)$$

where $X=[x_1\ x_2\ x_3\ x_4\ x_5\ x_6]^T$, the state vector, $U=[u_1\ u_2]^T$, the control vector and Q and R are the ponderation matrices [1, 2, 3].

$Q = \mathrm{diag}([0.2\ \ 2\ \ 0.2\ \ 8\ \ 4000\ \ 4000])$ and

$R = \mathrm{diag}([4\ \ 0.8])$.

3- Decomposition of the system:

In this work, the proposed method consists of the decomposition of the system into two subsystems. The first is a linear one, for which the optimal gain is obtained by solving the nonlinear Riccati equation. The second is also linear with respect to his state and control vector.. However, its dynamical behavior depends on the state vector of the first one

Consider the rotary crane described above (1). This system can be divided into two subsystems defined by equation (2) and (3) [1, 2, 3]. The subsystem 1 is defined by:

$$\begin{cases} x_{1,k+1} = x_{1,k} + T\, x_{2,k} \\ x_{2,k+1} = x_{2,k} + T\, u_{1,k} \end{cases} \quad (2)$$

The corresponding objective function to be optimized is described as follows:

$$J_1=\frac{1}{2}\sum_{i=1}^{200}(X_1^T Q_1 X_1 + U_1^T R_1 U_1)$$

where $Q_1 = \mathrm{diag}([0.2\ \ 2])$ and $R_1 = 4$

The subsystem 2 is described by:

$$\begin{cases} x_{3,k+1} = x_{3,k} + T\, x_{4,k} \\ x_{4,k+1} = x_{4,k} + T(u_{2,k} + \delta_1\, x_{5,k}) \\ x_{5,k+1} = x_{5,k} + T\, x_{6,k} \\ x_{6,k+1} = x_{6,k} - \dfrac{T}{(x_{1,k}+h)}(u_{2,k} + \delta_2\, x_{5,k} + 2\, x_{2,k}\, x_{6,k}) \end{cases} \quad (3)$$

We focus on the optimization of the quadratic criterion given below:

$$J_2=\frac{1}{2}\sum_{i=1}^{200}(X_2^T Q_2 X_2 + U_2^T R_2 U_2)$$

where $Q_2 = \mathrm{diag}([0.2\ \ 8\ \ 4000\ \ 4000])$ and $R_2 = 0.8$.

4- Optimal control:

4-1 Optimal control of subsystem 1:

Subsystem 1 being linear, also independent from subsystem 2, can be represented by the following condensed form of the state equation:

$$X_{1,k+1} = A_{1,k}\, X_{1,k} + B_{1,k}\, U_{1,k}$$

where:

$$A_1 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix},\ B_1 = \begin{bmatrix} 0 \\ T \end{bmatrix} \text{ and } X_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

The optimal solution is then expressed as:

$$U_{1,k} = -K_{1,k}\, X_{1,k}$$

$$K_{1,k} = (R_1 + B_{1,k}^T P_{1,k+1} B_{1,k})^{-1} B_{1,k}^T P_{1,k+1} A_{1,k}$$

where $P_{1,k}$ is the solution of the following Riccati equation

$$\begin{aligned} P_{1,k} = {} & Q_1 + A_{1,k}^T P_{1,k+1} A_{1,k} - \\ & A_{1,k}^T P_{1,k+1} B_{1,k} (R_1 + B_{1,k}^T P_{1,k+1} B_{1,k})^{-1} B_{1,k}^T P_{1,k+1} A_{1,k} \end{aligned} \quad (4)$$

Simulation results are given in figure 3.

4-2 Optimal fuzzy logic controller for subsystem 2:

Subsystem 2, whose state vector is $[x_3\ x_4\ x_5\ x_6]^T$, is linear for given values of x_1 and x_2. The main idea is to compute the optimal gain for different considered state point, in order to construct a fuzzy controller which generates the optimal control. The subsystem 2 is represented by the following condensed form:

$$X_{2,k+1} = A_{2,k}\, X_{2,k} + B_{2,k}\, U_{2,k} \qquad (5)$$

where:

$$A_2 = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & T\delta_1 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & -\alpha\delta_2 & -\alpha\beta \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ T \\ 0 \\ -\alpha \end{bmatrix}, X_2 = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}.$$

α and β are functions of x_1 and x_2:

$$\alpha = \frac{T}{x_{1,k} + h} \text{ and } \beta = 2x_{2,k}$$

The subsystem 2 depends on variables of state of the first subsystem. These parameters of interconnection that constitute variables x_1 and x_2 give back the control of the second subsystem difficult. So the direct approach considered constant, to every step of iteration, values taken by these variables of interconnection during the evolution of the system. We have gotten, in every iteration, a Riccati equation to solve to deduce the local optimal control. Simulation results are shown in figure 3. This method especially requires a relatively elevated time if we considered a more complex system.

The optimal solution, which minimizes the quadratic criterion J_2, is a function of x_1 and x_2. So, the optimal gain depends on x_1 and x_2. Then, one can easily write:

$$U_{2,k} = -K_{2,k}(x_{1,k}, x_{2,k}).X_{2,k} \qquad (6)$$

The main idea is to compute the optimal gain for different considered state point X_1, in order to construct a fuzzy controller which generates the optimal control [2, 3, 4].

The considered fuzzy controller is the Takagi-Sugeno's first order fuzzy system. The proposed system employs the symmetrical gaussian functions to represent the membership functions of the antecedent parts of fuzzy rules. The rules have the following form:

If (x_1 is F_1) and (x_2 is F_2) Then $u_2 = -(k_3x_3 - k_4x_4 - k_5x_5 - k_6x_6$.

For 100 given values of x1 and 100 given value of x2, we solve the Riccati equation. We obtain 10000 input-output pairs of data base. Then, a fuzzy logic system having two inputs (x_1, x_2) and four outputs (k_3, k_4, k_5, k_6)was constructed in order to identify these gains.

The structure of the fuzzy logic controller is given below:

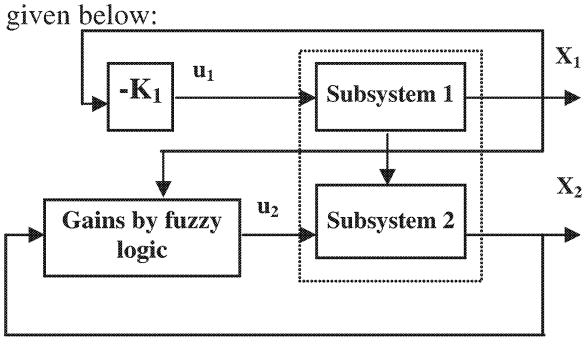


Figure 2: Structure of the fuzzy logic controller

The cost of the objective criterion are giving in table 1. Simulation results are shown in figure 3. We have used three gaussians membership functions for x_1 and three gaussians membership for x_2. It is clear that the state vector (x_1 and x_2) of subsystem 1 have rapid variation compared to the variation of the state vector of subsystem 2 (x_3 to x_6). This can be explained by the fact that the vertical movement of the load (subsystem 1) is practically independent from the swing (variables x_5 and x_6). However, the load horizontal movement (subsystem 2) is the principal factor generating the swing. Thus, the control vector of subsystem 2 is the principal factor responsible for the limitation of the hoisting. This is why subsystem 1 is more rapid than subsystem 2.

5- Robustness and stability [11, 12] :

For testing the robustness of this method, we have supposed a variation on the parameters of the system with random noises of null averages and maximum values 5%. Results simulations are shown in figure 4 when applying the direct method, and in figure 5 when applying the fuzzy logic controller.

We remark that the fuzzy controller ensure the robustness of the system when random noise affect the parameters of the system.

In order to study the stability of the system, firstly, we must prove the stability of each subsystem. The subsystem 1 is asymptotically stable if the matrix $F_1 = A_1 - B_1K_1$ is stable and if there is a common and definite symmetrical matrix positive P verifying: $F_1^T P F_1 - P < 0$,

However, with Matlab LMI Control Toolbox, it can be seen that the LMI is feasible and the solutions can be obtained as:

$$P_1 = \begin{bmatrix} 2.576 & 0.9795 \\ 0.9795 & 3.5986 \end{bmatrix}$$

The subsystem 2 is asymptotically stable if all the matrices $F_i = A_i - B_i K_i \; \forall i$ are stable and if there is a common and definite symmetrical matrix positive P verifying [12]:

$$F_{ii}^T P F_{ii} - P < 0 \,, \forall \; i \qquad (7)$$

$$\left(\frac{F_{ij} + F_{ji}}{2} \right)^T P + P \left(\frac{F_{ij} + F_{ji}}{2} \right) \leq 0 \,, \text{for } i < j \qquad (8)$$

With Matlab LMI Control Toolbox, it can be seen that the LMI is feasible and the solutions can be obtained as:

$$P_2 = \begin{bmatrix} 0.0623 & 0.22 & -3.8147 & 1.4002 \\ 0.22 & 2.1461 & -37.2106 & 12.4141 \\ -3.8147 & -37.2106 & 927.5028 & -58.0158 \\ 1.4002 & 12.4141 & -58.0158 & 310.6831 \end{bmatrix}$$

We suppose that $P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$ which verify the Lyapunov condition. So, we conclude that the system is stable.

6- Conclusion:

This paper concerns the determination of the optimal control strategies of a rotary crane. The system was decomposed into two linear interconnected subsystems. Dynamics of each subsystem depend on the state vector of the other one. In this context, linear feedback gains are considered. The optimal gains of each subsystem depend on the state of the other one. Takagi and Sugeno's fuzzy models and some stability results have been presented. Simulation results show satisfactory performances.

References:

[1] C. Rekik, M. Djemel, N. Derbel: "Neural Network Approach for the Optimal Control of a Container Crane". Middle East Symp. on Simulation and modeling, MESM'02, Jordan, Amman, Septmeber 2002.

[2] C. Rekik, M. Djemel, N. Derbel, A. Alimi: "Design of Optimal Fuzzy Logic Controller with Genetic Algorithms". International Symp. Intelligent Control, ISIC'02, Canada, october 2002.

[3] A. Ouezri, N. Derbel: "On the Improvement of the Dynamical Behavior of a Rotary Crane Optimal Control, Neuro-Fuzzy Approach". Int. Conf. Advances Mechanical Engineering (ICAME'02), Hammamet, Tunisia, March 2002.

[4] N. Derbel: "On the Optimal Control Approaches of Nonlinear Systems", Int. Conf. on Artificial Computational on Decision, Control and Application, ACIDCA'00, Monastir-Tunisia, March 2000

[5] J. S. Jang, Self-learning fuzzy controller based on temporal back propagation, IEEE Tans Neural Networks. . vol. 5, (1990) 714-723.

[6] J. S. Jang, ANFIS adaptive-network-based fuzzy inference system. IEEE Trans. Systems Man Cybern. vol. 3 (1993) 665-683.

[7] H.K. Kwan and Y. Cai, A fuzzy Neural Network and its application to pattern recognition. IEEE Trans. Fuzzy System vol. 2, no 3 (1994) 185-193.

[8] J. Yen, R. Langari and L. Zadeh: "Industrial Applications of Fuzzy Logic and Intelligent Systems". New York, IEEE Press, 1995.

[9] Y. Sakawa and Y. Shindo: "Optimal control of container cranes". Automatica, vol. 18, pp. 257-266, 1981.

[10] Y. Sakawa, Y. Shindo and Y. Hashimoto: "Optimal control of rotary crane", Journal of optimization theory and applications, vol. 35, No. 4, December 1981.

[11] T. Takagi and M. Sugeno: "Fuzzy identification of systems and its applications to modelling and control". IEEE Trans. on SMC, vol. 15, no. 1, pp. 116-132. 1985

[12] K. Tanaka T. Ikeda and H.O. Wang (1998): "Fuzzy regulators and fuzzy observers: relaxed stability conditions and LMI-based designs". IEEE Trans. fuzzy systems, vol. 6, no. 2 pp. 251-256.

Table 1: Cost evaluation for different approaches

Approaches	DDP	Direct approach	Fuzzy logic controller
Criterion (J)	258.1664	258.2878	258.2137

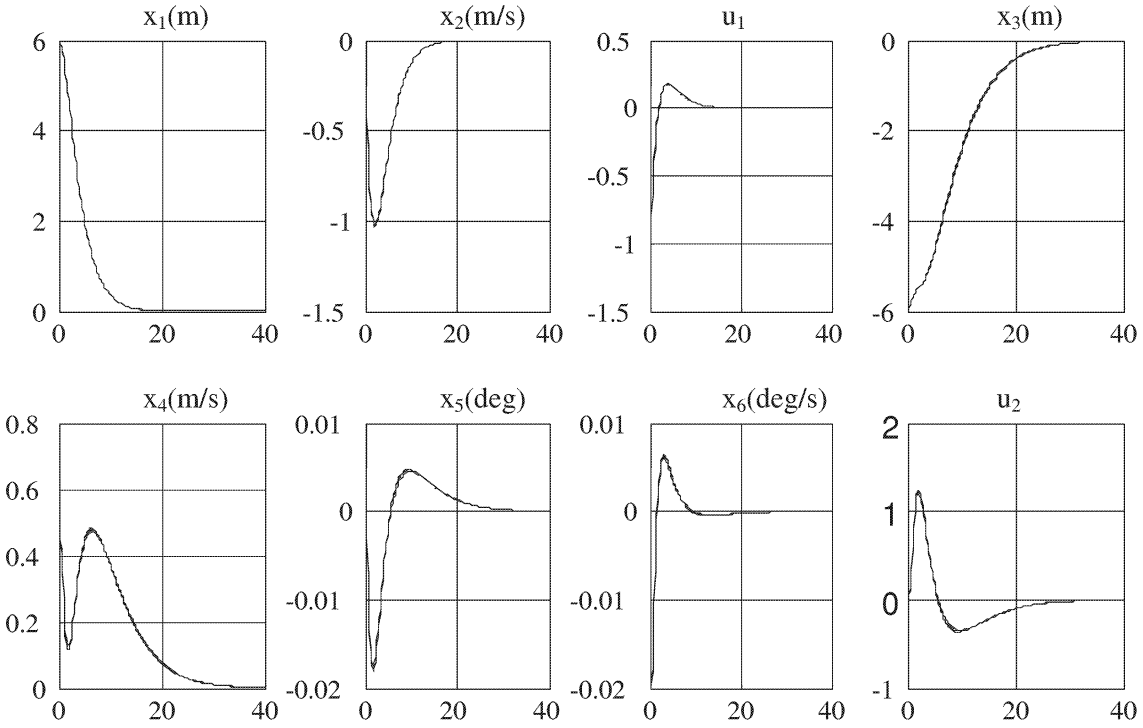


Figure 3: Simulation results:
Dotted line: Classical method (DDP).
Dashed line: Direct approach.
Continued line: Fuzzy logic controller.

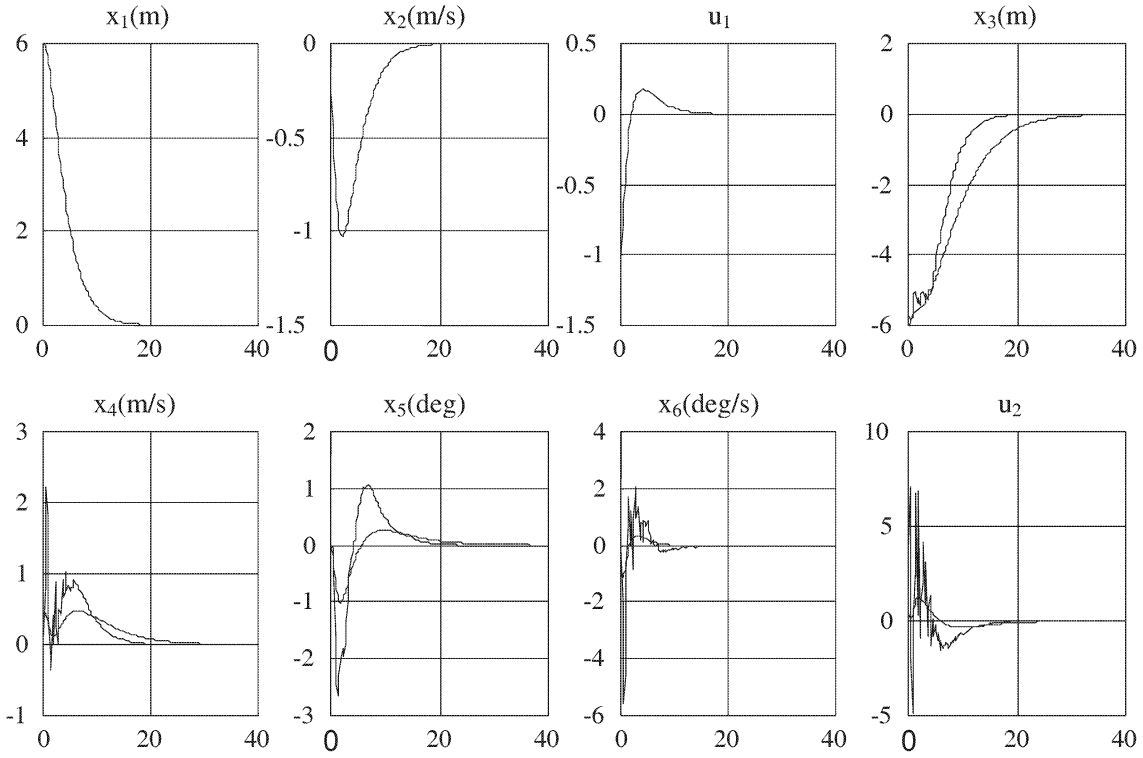


Figure 4: Simulation results applying direct approach :
Dotted line: not disturbed system.
Continued line: disturbed system.

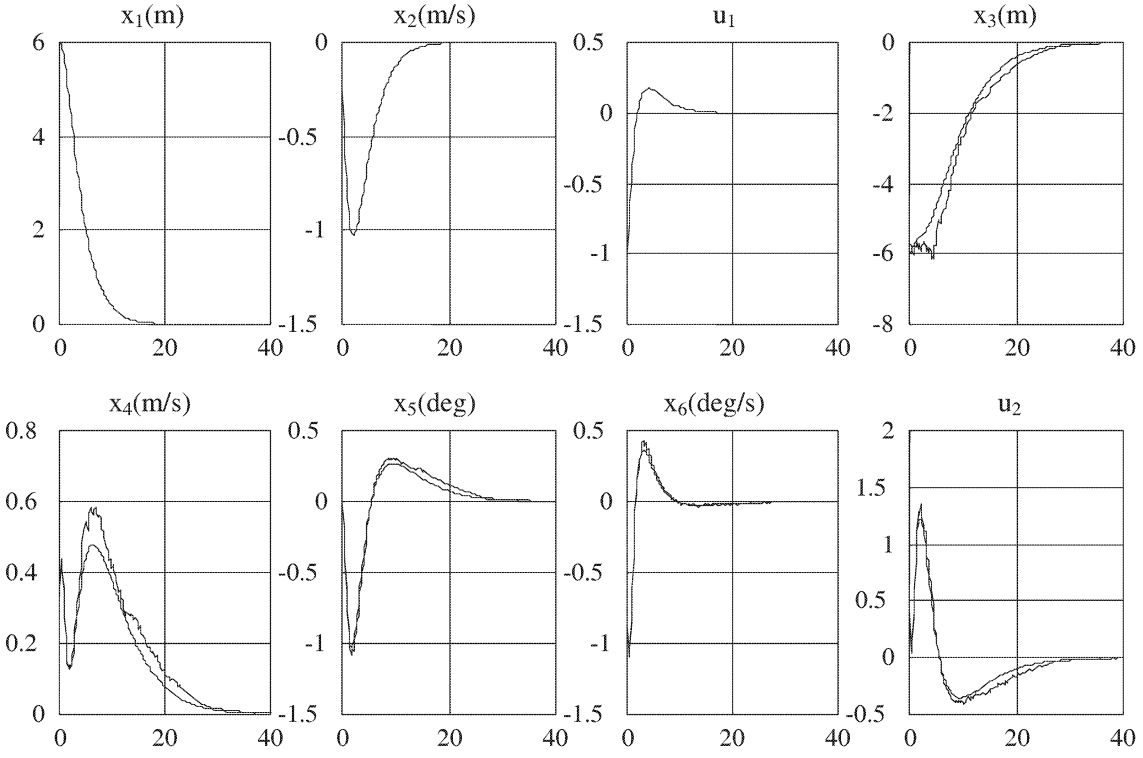


Figure 5: Simulation results applying fuzzy logic controller:
Dotted line: not disturbed system.
Continued line: disturbed system.

SIMULATION OF PID CONTROLLERS

On the Particularities of Hysteresis Band Widths in Takahashi DTC Strategy: the So-called Focal Speeds

Fatma Ben Salem and Ahmed Masmoudi
Research Unit on Renewable Energies and Electric Vehicles (RELEV)
University of Sfax, Sfax Engineering School, BP W, 3038 Sfax, Tunisia
e-mails: fatma.bensalem@enis.rnu.tn, a.masmoudi@enis.rnu.tn

Abstract– **Hysteresis band widths of both torque and flux regulators greatly influence the performance of DTC strategies. The paper deals with the influence of these bands on the inverter average switching frequency, which is the result of the one due to the torque regulator and of the one due to the flux regulator. Of particular interest are the so-called focal speeds for which the inverter average switching frequency due to the torque regulator is null. A special attention is focused upon such speeds which are derived in the case of a constant load torque. Simulation results have confirmed analytical predictions.**

I. INTRODUCTION

In recent years, DTC has been presented as an alternative control method to Field Oriented Control (FOC) [1]. The popularity of DTC has been gained thanks to several advantages, such as: higher torque dynamics, high robustness against the variations of the drive parameters, appropriateness to sensorless control, etc [2, 3].

DTC was introduced empirically by *Takahashi* in the mid 1980s [3]. The idea consists of describing the way in which the torque and the stator flux are directly controlled, without the intermediate step of using current regulators as in field oriented control. Basically, the *Takahashi* DTC strategy selects one of the inverter's six voltage vectors in order to keep the stator flux and torque within given hysteresis bands.

Since the proposal of *Takahashi*, many works have been focused towards the analytical basis of DTC. Within this trend, different approaches have proposed to give a firm explanation of the *Takahashi* DTC strategies, and to improve it. Of particular interest are: (i) the introduction of two lock-up tables taking into account the torque variation with respect to time [4, 5], (ii) the reduction of the flux ripple by considering twelve sectors instead of six [6], and (iii) the improvement of the DTC performance by selecting appropriate hysteresis controllers for the flux and torque loops [7].

Within the last item, it has been noted that the band width of the hysteresis controllers strongly influences the drive performances, especially the uncontrolled current harmonics which are associated with high torque ripples and high inverter switching losses [8, 9, 10, 11]. Even though hysteresis bands are kept constant, it has been found in [12, 13] that the inverter switching frequency varies according to the operating conditions such as the motor speed.

The present paper is devoted to an analysis of the effects of the hysteresis bands of the flux and torque regulators on the inverter switching frequency in the *Takahashi* DTC strategy. A special attention is focused upon the so-called focal speeds for which the inverter average switching frequency due to the torque regulator is null. Analytical prediction of the expressions of the focal speeds is considered in the case of a constant load torque.

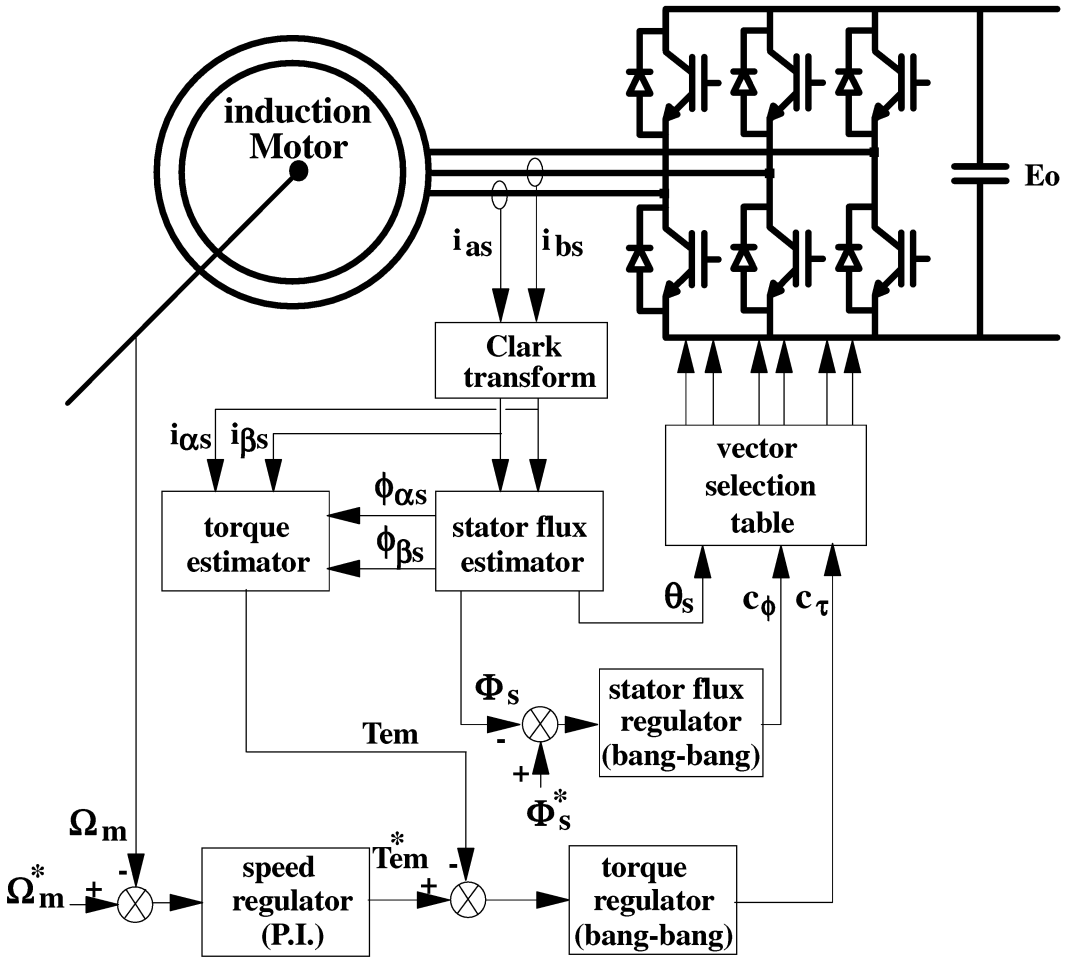


Figure 1: DTC Block Diagram (*Takahashi* approach).

II. PRINCIPLE OF THE TAKAHASHI DTC STRATEGY

The basic DTC strategy, to which is referred most if not all the literature in this area, is the one developed in 1986 by *Takahashi* [3].

The idea of the *Takahashi* DTC strategy, whose block diagram is shown in figure 1, is to choose the best vector of voltage which makes the flux rotate and produce the desired torque.

Considering a three-phase two-level voltage source inverter, there are six non-zero voltage vectors and two zero voltage vectors which can be applied to the machine terminals.

The set of the combinations of the voltage vector voltages proposed by *Takahashi* is provided in table I [3].

Tabla I

Basic DTC switching table.

c_ϕ	+1			-1		
c_τ	+1	0	-1	+1	0	-1
S_1	V_2	V_7	V_6	V_3	V_0	V_5
S_2	V_3	V_0	V_1	V_4	V_7	V_6
S_3	V_4	V_7	V_2	V_5	V_0	V_1
S_4	V_5	V_0	V_3	V_6	V_7	V_2
S_5	V_6	V_7	V_4	V_1	V_0	V_3
S_6	V_1	V_0	V_5	V_2	V_7	V_4

The stator flux vector can be estimated using measured current and voltage vectors:

$$\frac{d}{dt}\overline{\Phi}_s = \overline{V}_s - R_s\overline{I}_s \quad (1)$$

The electromagnetic torque can be expressed in terms of stator current and flux as:

$$T_{em} = N_p Im\left(\overline{\Phi}_s\overline{I}_s^*\right) \quad (2)$$

where N_p is the pole pair.

An other expression of the T_{em} has been considered in the literature [14]:

$$T_{em} = N_p\frac{M}{l_r l_s - M^2}\|\overline{\Phi}_s\|\ \|\overline{\Phi}'_r\|\sin\delta \quad (3)$$

where Φ'_r is the rotor flux referred to the stator winding, δ is the angle between the stator and rotor flux-linkage space vectors, l_s, l_r and M are the stator self inductance, the rotor self inductance and the mutual inductance, respectively.

the average torque switching frequency is expressed as follows:

$$\begin{cases} S_a^+ = -\frac{R_{sr}}{\sigma l_s}T_{em} + \frac{N_p}{\sigma l_s}\left[\frac{3E_0}{\sqrt{2}\pi}\Phi_s - N_p\Omega_m\Phi_s^2\right] \\ S_a^- = -\frac{R_{sr}}{\sigma l_s}T_{em} - \frac{N_p^2}{\sigma l_s}\Omega_m\Phi_s^2 \end{cases} \quad (7)$$

where E_0 is the DC link voltage.

The average torque switching frequency is expressed as follows:

$$F_\tau = 2\frac{1}{T_r + T_f} = \frac{S_a^+ S_a^-}{\varepsilon_\tau\left(S_a^- - S_a^+\right)} \quad (8)$$

where T_r is the torque rising time, T_f is the torque falling time and ε_τ is the hysteresis band width of the torque controller.

Accounting for equations (7), the expression of F_τ given by (8) turns to be:

$$F_\tau = \frac{\left\{\frac{R_{sr}}{\sigma l_s}T_{em} - \frac{N_p^2}{\sigma l_s}\left[\frac{3E_0}{\sqrt{2}\pi N_p}\Phi_s - \Omega_m\Phi_s^2\right]\right\}\left\{\frac{R_{sr}}{\sigma l_s}T_{em} + \frac{N_p^2}{\sigma l_s}\Omega_m\Phi_s^2\right\}}{\varepsilon_\tau\ \frac{N_p}{\sigma l_s}\frac{3E_0}{\sqrt{2}\pi}\Phi_s} \quad (9)$$

III. AVERAGE INVERTER SWITCHING FREQUENCY: FORMULATION

The width of the torque and flux hysteresis bands affects the inverter switching frequency and then the switching losses [15, 16]. The average inverter frequency F_c is the result of the one due to the torque hysteresis controller F_τ and of the one due to the flux hysteresis controller F_ϕ, such that:

$$F_c = F_\tau + F_\phi \quad (4)$$

Assuming that under steady state, the torque ripples do not exceed the limits of the bandwidth of the hysteresis, and if the leakage flux is neglected, the positive slope is given as follows [12]:

$$S^+ = -\frac{R_{sr}}{\sigma l_s}T_{em} + \frac{N_p}{\sigma l_s}\left[Im\left(\overline{V}_s\overline{\Phi}_s\right) - N_p\Omega_m\overline{\Phi}_s^2\right] \quad (5)$$

The negative slope is obtained by setting $\overline{V}_s = 0$:

$$S^- = -\frac{R_{sr}}{\sigma l_s}T_{em} - \frac{N_p^2}{\sigma l_s}\Omega_m\overline{\Phi}_s^2 \quad (6)$$

where $R_{sr} = r_s + \frac{l_s}{l_r}r_r$ and Ω_m is the speed.

The study carried out in to [15, 16] proves that the slopes of the stator flux magnitude are highly nonlinear functions. Thus the calculation of the average switching frequency F_ϕ of the flux hysteresis controller turns to be different from (8).

The average flux switching frequency F_ϕ can be obtained as [12]:

$$F_\phi = N_\phi f_s \quad (10)$$

where N_ϕ is the number of the changes of voltage vector per cycle and f_s is the stator frequency.

It is to be noted that equation (9) varies with the speed.

Of particular interest are the values of the speed for which F_τ is equal to zero: the so-called focal speeds Ω_{mf}. These depend on the load torque expression. In what follows, the expressions of the focal speeds in the case of a constant load torque are derived.

IV. FOCAL SPEED DERIVATION

Let us consider the case of a constant load torque $T_l = K_0$, the expressions of the speeds for which $F_\tau = 0$ are the following:

$$\begin{cases} \Omega_1 &= -\dfrac{K_0 R_{sr}}{N_p^2 \Phi_s^2} \\ \Omega_2 &= \dfrac{3E_0}{\sqrt{2}\pi N_p \Phi_s} - \dfrac{K_0 R_{sr}}{N_p^2 \Phi_s^2} \end{cases} \quad (11)$$

It is to be noted that Ω_2 is takes positive values, while Ω_1 is negative and has to be discarded as far as it corresponds to generator operation.

In the particular case of no-load operation, equations (11) are reduced to:

$$\begin{cases} \Omega_{01} &= 0 \\ \Omega_{02} &= \dfrac{3E_0}{\sqrt{2}\pi N_p \Phi_s} \end{cases} \quad (12)$$

The expressions of the focal speeds for different types of the load torque have been established in [13].

V. CASE STUDY

The parameters of the induction motor considered in this section are listed in table II. It has the following ratings: 220V, 10kW and 1470rpm at 50Hz.

Table II: Induction motor parameters

$r_s = 0.29\Omega$	$r_r = 0.38\Omega$	$M = 47.3$mH
$l_s = l_r = 50$mH	$N_p = 2$	$J = 0.5$Kg.m^2

A. Analysis Based on the Analytical Study

Figures 2 and 3 show the variations of F_τ with respect to the Ω_m for different values of ε_τ (from 2% to 10% step 2%), in the case of a constant nominal load torque $T_{ln} = 63.7$Nm and in the case of no-load operation, respectively.

These figures highlight the existence of the focal speeds. In the case of the constant nominal load torque $T_{ln} = 63.7$Nm (figure 2), the values of the focal speeds are: $\Omega_1 \simeq$-102rpm and Ω_2=1402rpm. In the case of no-load operation (figure 3), Ω_{01}=0rpm and Ω_{02}=1504rpm.

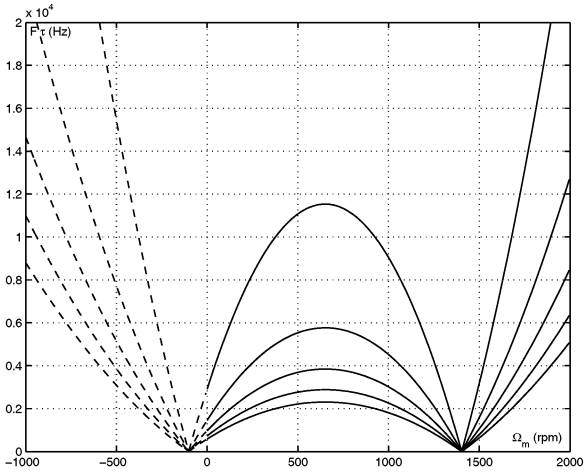


Figure 2: F_τ versus Ω_m for ε=10%, 8%, 6%, 4% and 2% (bottom to top), for $T_l = T_{ln} = 63.7$Nm

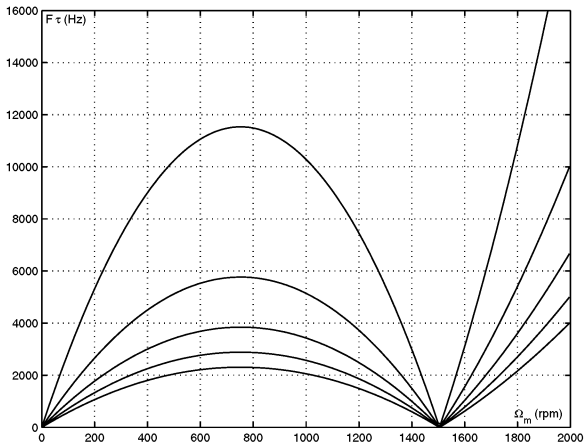


Figure 3: F_τ versus Ω_m for ε=10%, 8%, 6%, 4% and 2% (bottom to top), at no-load operation

B. Simulation Results

Figures 4.a, 4.b and 4.c show T_{em}, Φ_s and F_c, respectively, versus time during the start-up of the induction motor under DTC and at lo-load operation. During this start-up, the motor follows a reference speed which varies linearly with respect to time from 0 to 1600rpm during 5s. One can notice the low values of F_c for low speeds and for a speed which is almost equal to Ω_{02}.

Figures 5 and 6 show F_c versus time at steady state operation corresponding to a speed equal to the focal speed Ω_{02}, for ε_ϕ=2% and 10%, and for ε_τ=5% and 10%, respectively. One can notice that F_c depends on ε_ϕ while it is independant of ε_τ, which clearly confirm the analytical results.

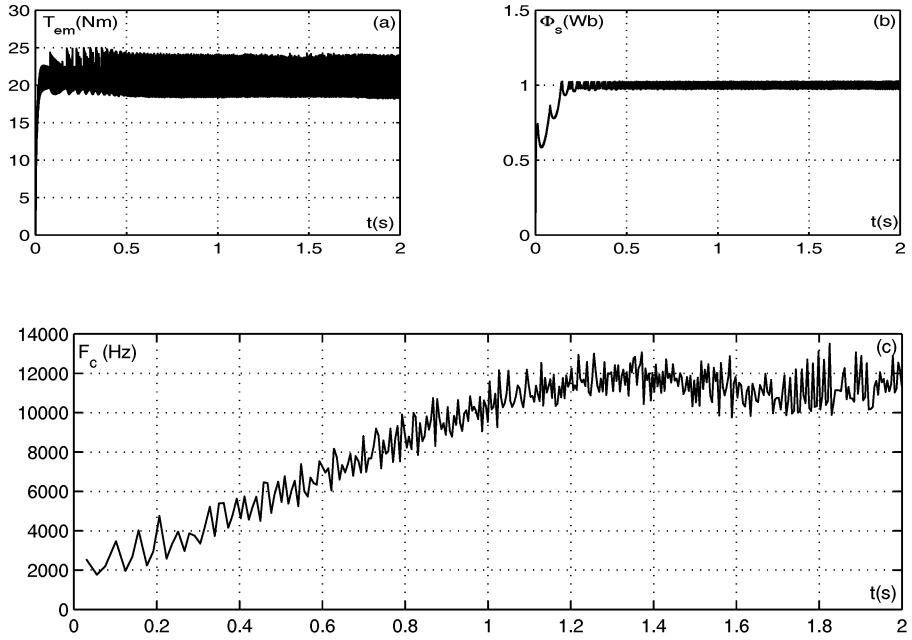


Figure 4: Dynamic of the induction motor under DTC and at lo-load operation during the start-up. (a): electromagnetic torque T_{em}, (b): stator flux Φ_s and (c): inverter average switching frequency F_c

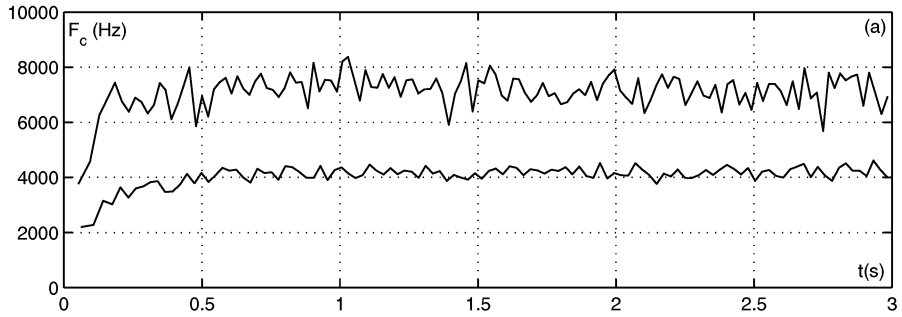


Figure 5: F_c versus time at the focal speed Ω_{02} and for ε_ϕ=2% (up) and 10% (bottom)

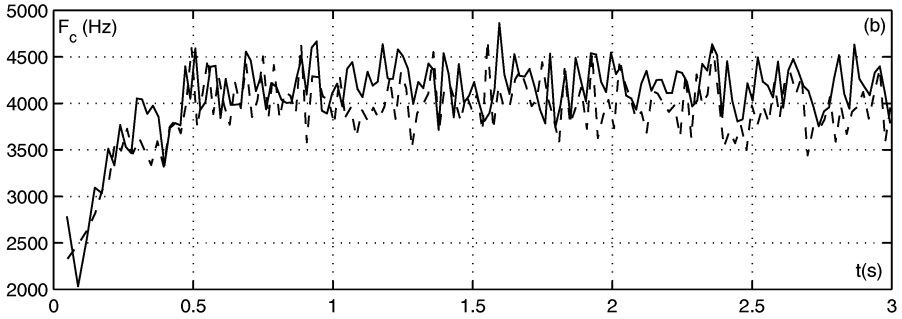


Figure 6: F_c versus time at the focal speed Ω_{02} and for ε_τ=5% (dashed) and 10% (continuous)

VI. CONCLUSION

This paper treated the effects of both torque and flux hysteresis bands on the inverter average switching frequency. This is has been done considering an analytical prediction of the switching frequency due to the torque regulator and the one due to the flux regulator.

It has been found that the flux switching frequency is a linear function of the speed, while the torque switching frequency is a nonlinear function of the speed. Of particular interest are the so-called focal speeds for which the torque switching frequency is null. These have been derived in the case of a constant load torque. Simulation results have confirmed analytical predictions.

REFERENCES

[1] F. Blaschke, "The principal of field oriented applied to the new transvector closed loop control system for rotating field machines," *Siemens Revue*, vol. 39, pp. 217-220,1972.

[2] J. N. Nash, "Direct torque control, induction motor vector control without an encoder," *IEEE Trans. on Industry Applications*, vol. 33, pp. 333-341, 1997.

[3] I. Takahashi and T. Noguchi, "A new quick-response and high-efficiency control strategy of an induction motor," *IEEE Trans. on Industry Applications*, vol. IA-16, pp. 820-827, 1986.

[4] I. El Hassan, X. Roboam, B. De Fornel, and E. V. Westerholt, "Torque dynamic behavior of induction machine direct torque control in four quadrant operation," *Proc. IEEE ISIE'97 Conference*, pp. 1034-1038, Lisboa, Portugal, 1997.

[5] F. Ben Salem, A. Yangui and A. Masmoudi, "Induction motor direct torque control: a comparison between two strategies," CD-ROM *Third Int. Conf. on Systems, Signals and Devices* (SSD'03), Sousse, Tunisia, 2003.

[6] Y. H. Liu, C. L. Chen, and R. J. Tu, "A novel space vector currentregulation scheme for a field-oriented controlled induction motor drive," *Int. Conf. PCIM Nrberg/Germany*, 1998.

[7] C. Lascu, I. Boldea, and F. Blaabjerg, "A modified direct torque control for induction motor sensorless drive," *IEEE Trans. Industry Applications*, vol. 36, pp. 122-130, 2000.

[8] D. Casadei, G. Grandi and A. Tani, "Effect of flux and torque hysteresis band amplitude in direct torque control of induction machines," *Proc. IEEE IECON'94*, pp. 299-304, 1994.

[9] D. Casadei, G. Grandi and A. Tani, "Switching strategies in direct torque control of induction machines," *Proc. Int. Conf. on Electrical Machines*, vol. 2, pp. 204-209, Paris, France, 1994.

[10] J.K. Kang and S.K. Sul, "New direct torque control of induction motor for minimum torque ripple and constant switching frequency," *IEEE Trans. Industry Applications*, vol. 35, pp. 1076-1082, 1999.

[11] C. Lochot, X. Roboam, and P. Maussion, "A new direct torque control strategy for an induction motor with constant switching frequency operation," *Proc. EPE'95*, pp. 2-431-2.436, Seville, spain, 1995.

[12] F. Ben Salem and A. Masmoudi, "Effects of hysteresis bands, the speed range and the sampling time on the inverter switching frequency in a DTC strategy (Part 1)," CD-ROM *Third Int. Conf. on Systems, Signals and Devices* (SSD'05), Sousse, Tunisia, 2005.

[13] F. Ben Salem and A. Masmoudi, "Effects of hysteresis bands, the speed range and the sampling time on the inverter switching frequency in a DTC strategy (Part 2)," CD-ROM *Third Int. Conf. on Systems, Signals and Devices* (SSD'05), Sousse, Tunisia, 2005.

[14] T. G. Habetler, F. Profumo, M. Pastorelli and L. M. Tolbert, "Direct torque control of induction machines using space vector modulation," *IEEE Trans. on Industry Applications*, vol. IA-28, no. 5, pp. 1045-1053, 1992.

[15] J.-K. Kang and S.-Ki Sul, "Analysis and prediction of inverter switching frequency in direct torque control of induction machine based on hysteresis bands and machine parameters," *IEEE Trans. on Industrial Electronics*, vol. 48, no.3, pp. 545-553, 2001.

[16] J.-K. Kang, D.-W. Chung and S.-K. Sul, "Direct torque control of induction machine with variable amplitude control of flux and torque hysteresis bands," *Conf. Rec. IEEE-IAS*, pp. 640-642, 1999.

PID Control Design of A Blower System using the pole compensation technique

Houda Ben Jmaa Derbel[1,2]

[1]: Research Unit on Renewable energies and ELEctric Vehicles (RELEV)
Ecole Nationale d'Ingénieurs de Sfax, 3038, Sfax, Tunisie
[2]: Faculté des Sciences de Sfax, 3038, Sfax, Tunisie
Email. h.derbel@fss.rnu.tn

Abstract: *This paper deals with the PID control of a blower system which is modelled by a linear second order model with pure delay and uncertain static gain. Time constants of the PID controller are determined by compensating the poles of the model. The PID proportional gain is determined with the minimization of the settling time by the use of genetic algorithms as tools of optimization. The obtained PID parameters give good and satisfactory performances. A discussion with respect to the operating point is also presented.*

Keywords: *PID Control, Delay Systems, Genetic Algorithms, Parameter uncertainties, Blower System.*

1. Introduction

In this work, we are applying linear techniques to a nonlinear problem. The subject of the application of linear methods to control nonlinear perturbations of dynamic systems has been the subject of intense research effort over a long number of years, and there is a very extensive literature available [5, 6, 13, 14]. It is in this area that the major contribution of this paper lie.

The objective of this work is the determination of PID controller parameters of a blower system. It is well known that the dynamical behavior of such system can be described by a complex mathematical model containing partial derivative terms of the temperature with respect to space variables. Controlling this flow is however still an extremely challenging problem. To overcome such difficulty, an uncertain linear reduced model is considered. Such model is a second order one with a pure delay time.

In the literature, PID controllers [2, 3, 4] were extensively studied in different research area and for different application area (tuning PID [1, 7, 8, 18, 19], optimal PID tuning [15, 16, 17], etc.)

In this paper, parameters of the PID controller are then determined by compensating the model poles and by minimizing the settling time using genetic algorithms. A stability of the closed loop system is then studied, taking into account the variations of the static gain. Simulation results shows that the obtained performances are slightly affected by the system gain variations.

This paper is organized as follows. Section two presents the problem formulation of the blower system. Section three is dedicated to the design of PID controllers of delayed second order models. Section four presents some simulation results illustrating the proposed approach.

2. Problem Formulation

The considered blower is a cylindrical volume whose axis is parallel to the (ox) direction, having a section S and a lateral thickness e. The fluid is heated by a power heating Q which is placed below a fan that creates the fluid flow (see figure 1).

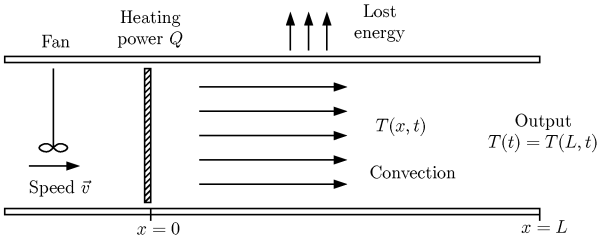


Figure 1: Schematic representation of the cylindrical blower system.

The dynamical behavior of the temperature inside the blower can be described by the heat bal-

ance expressed by the following differential equation [9, 10, 11, 12]:

$$\rho C_p \frac{\partial T}{\partial t} = -\rho C_p v \frac{\partial T}{\partial x} - \frac{2}{R} \times \frac{T - T_a}{\frac{1}{h_i} + \frac{1}{h_e} + \frac{e}{\lambda_w}} \quad (1)$$

where ρ is the air density, C_p is the air specific heat, e is the thickness of lateral sides (wall), R is the blower radius, h_i and h_e are respectively the internal and the external convective heat coefficient and λ_w is the blower wall conductivity. v is the air speed inside the blower and T_a represents the ambient temperature. The boundary conditions are the heating power Q and the ambient temperature at the coordinate $x = 0$. The heat exchange through the lateral wall is assumed as the resultant heat transferred by the interior convection, the conduction and the exterior convection.

The following are the blower parameters: $R = 0.125m$, $e = 2cm$, $L = 1.2m$ is the blower length; $v_a = 0.1ms^{-1}$ the air speed outside the blower, $v = 3ms^{-1}$, $0 \leq T_a \leq 27°C$, and $0 \leq Q \leq 5kW$.

Equation (1) contains partial derivatives of the temperature inside the blower with respect to time and space. By solving this equation and observing the dynamical response of the output temperature of the blower, it is easy to show that system can be modelled by a linear model with a pure delay. The choice of a second order model with a pure delay is adequate [20, 21].

$$G(p) = \frac{ke^{-\tau p}}{1 + 2\xi\frac{p}{\omega_n} + \left(\frac{p}{\omega_n}\right)^2} \quad (2)$$

The optimization of parameters k, τ and T by genetic algorithms gives: $\tau = 11.5$, $\xi = 0.8$, $\omega_n = 0.8$ and $170 \leq k \leq 340$ [10, 11, 12]. Parameter k, which is uncertain in this model, depends on the operation point around which the system operates. In the following, we will consider $k = k_{mean} = 255$.

3. PID Control of a Delayed Second Order Model

Many works has been devoted to the synthesis of PID parameters in order to control delayed systems. In this context, we consider a linear second order system with a pure delay described by system (2).

The PID controller has the following transfer function:

$$C(p) = k_p \left(1 + \frac{1}{T_i p} + T_d p\right) \quad (3)$$

Time constants of the PID controller (T_i and T_d) are determined by compensating the poles of the system. This gives:

$$T_i = \frac{2\xi}{\omega_n} = 2 \quad (4)$$

$$T_d = \frac{1}{2\xi\omega_n} = 0.7813 \quad (5)$$

It is to be noted that the compensation technique can be used to eliminate complex poles if and only if these poles are sufficiently damped, that is to say if their damping factors are larger than 0.5. In our application, the damping factor of the two complex poles is equal to 0.8.

The open loop transfer function of the controlled system becomes:

$$OLTF(p) = \frac{kk_p}{T_i p} e^{-\tau p} \quad (6)$$

The gain k_p is determined in such away the closed loop system has the minimum settling time τ_s defined by:

$$\forall t \geq \tau_s, \ \frac{|s(t) - s(\infty)|}{|s(0) - s(\infty)|} \leq 0.05 \quad (7)$$

Optimizing k_p by genetic algorithms gives the best results for:

$$k_p = \frac{0.53}{k} \times \frac{T_i}{\tau} \simeq \frac{1}{2k} \times \frac{T_i}{\tau} \quad (8)$$

It is notable that results yielded from genetic algorithms give $k_p = 0.53T_i/(k\tau)$. This leads to an overshoot equal to 5%. The approximation given by $k_p = T_i/(2k\tau)$ gives an overshoot less than 4.1% yielding a very small increase on the settling time.

The chosen parameters of the PID controller correspond to a controlled system yielding damped oscillations, with an overshoot equal to 4% and a settling time expressed by:

$$t_s = 3.36\tau \quad (9)$$

It is clear that this result depends only on the delay τ. This can be explained by the fact that the characteristic equation is independent on parameters T and k:

$$1 + C(p)G(p) = 0$$

or:

$$1 + \frac{k\ k_p}{T_i p} e^{-\tau p} \quad = \quad 0$$

or well:

$$1 + \frac{1}{2\ \tau\ p} e^{-\tau p} \quad = \quad 0 \qquad (10)$$

In this case, the closed loop system is stable, having a phase margin equal to $90(1 - 1/\pi) \simeq 61.4°$ and a gain margin equal to π or 22.9 dB.

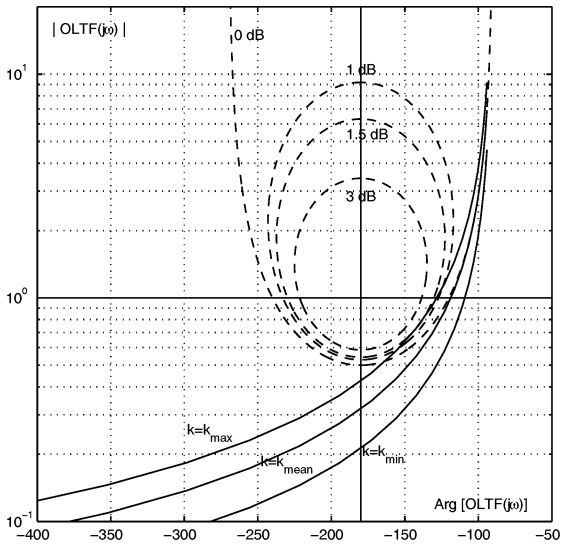


Figure 2: A discussion on the stability study in the Black diagram, for different values of the static gain of the model.

In order to check the stability of the system when the static gain varies, figure 2 shows the Black diagram of the closed loop system versus the static gain variations. Based on this figure, one can easily notice that the closed loop system is stable. Different stability margins are presented in Table 1. This Table shows that the system has a phase margin larger than 50°, a gain margin larger than 17 dB and a resonance factor smaller than 1 dB. Moreover, it is obvious that if the gain increases, the system response presents damped oscillations.

Table 1: Evolution of the phase margin M_φ, the gain margin M_G and the resonance factor Q_r, in terms of the static gain.

k	M_φ (°)	M_G (dB)	Q_r (dB)
k_{min}	70.8	30.9	0
k_{mean}	61.2	22.9	0
k_{max}	51.5	17.0	1.5

4. Application to the blower system

In the following, we will apply the determined PID controller to the blower system. The evolution of the output temperature of the blower is presented in figure 3. In this figure, the evolution of the output of the model is also presented. The two responses are quite similar. However, it is obvious that there is a static error between the control of the blower system (power heating Q) and the control of its model. This error can be justified by the variations of the uncertain static gain. The analysis of these two responses gives the following information:

- the settling time of the blower system is equal to 43 s, however the one of its model is equal to 38.5 s,
- the overshoot of the response of the blower system is equal to 4.81 %, however the one of its model is equal to 4.13 %.

In order to observe the dynamical behavior of the blower system for different operating points, figure 4 gives the evolution of the output temperature and the power heating of the blower system for different desired temperatures. It is obvious to conclude that the system presents damped oscillations for large values of the desired temperature. Consequently, the overshoot increases when the desired temperature increases. This can be explained by the fact that the static gain increases when the temperature increases. It is to notice that the minimum settling time is obtained for a desired temperature around 600 K for which the static gain is in its mean value. However, the system becomes more damped for small temperatures, because the static gain is smaller. In contrast, the system presents damped oscillations whose amplitude increases when the temperature increases, because, the static gain becomes larger. These results are confirmed by Table 2 which gives the evolution of the settling time and of the overshoot in terms of the desired operating point.

Table 2: Evolution of the settling time τ_s and the overshoot O_s, in terms of the desired temperature T_d.

T_d (K)	400	500	600	700	800
τ_s (s)	58.3	48.7	38.5	63	64
O_s (%)	0.5	1.9	4.8	8.6	12.6

5. Conclusion

This paper considered the PID control of systems with pure delays. In this context, systems will be modelled by a second order model with pure delays. The optimization phase uses genetic algorithms for the determination of the model parameters. On the basis of this model, time constants of the PID controller were determined by the pole compensation technique. The PID gain was determined by the minimization of the settling time, using genetic algorithms as optimization tools. The obtained PID parameters give good and satisfactory performances. A discussion with respect to the static gain of the model showed the stability and the robustness of the closed loop control system. A discussion with respect to the operating point showed that system presents damped oscillations for which the amplitude of oscillations increase for high temperature. This was justified by the fact that the system presents high static gain when the temperature increases.

References

[1] R. Argelaguet, M. Pons, . Aguilar-Martin and J. Quevedo: *"A New Tuning of PID Controllers based on LQR Optimization"*. European Control Conference, Karlsruhe, Germany, September 1997.

[2] K. J. Aström and P. Albertos: *"PID Control"*. Control Engineering Practice, vol. 9, pp. 1159-1161, 2001.

[3] K. J. Aström and T. Hagglund: *"The future of PID Control"*. Engineering Practice, vol. 9, pp. 1163-1175, 2001.

[4] K. J. Aström and T. Hagglund: *"PID Controller, Theory, Design and Tuning"*. Instrument Society of America, Research Triangle Park, NY, 1995.

[5] Baramov L., Tutty O. R. and Rogers E.: *"H_∞ Control of Nonperiodic Two-Dimensional Channel Flow"*. IEEE Trans. Control Systems Technology, vol. 12, no. 1, pp. 111-122, January 2004.

[6] Baramov L., Tutty O. R. and Rogers E.: *"Robust Control of linearized Poiseuille Flow"*. AIAA Journal, Guid. Dynamic Control, vol. 25, pp. 145-151, 2002.

[7] G. Calcev and R. Gorez: *"Iterative Technique for PID Controller Tuning"*. IEEE Conference on Decision and Control, CDC'34, New Orleans, 1995.

[8] W. D. Chang, R. C. wang and J. G. Hsieh: *"A Self Tuning PID Control for a Class of Nonlinear Systems Based on the Lyapunov Approach"*. Journal of Process Control, vol. 12, pp. 233-242, 2002.

[9] H. Ben Jmaa Derbel H.: *"On the Optimal PID Regulator for the Control of a Blower System"*. Accepted for publication in Asian Journal of Information Technology, (To appear 2005).

[10] H. Ben Jmaa Derbel H.: *"On the PID Control of a Temperature Regulation Process, Optimisation by Genetic Algorithms"*. Middle East Symposium on Modelling, MESM'03, Sharjah-UAE, January 2004.

[11] H. Ben Jmaa Derbel H.: *"Optimal Regulation Control of a Blower System"*. CATAEE'04, Amman-Jordan, Marsh 2004.

[12] H. Ben Jmaa Derbel H.: *'Optimisation par Algorithmes Génétiques d'un Régulateur PID pour la Commande d'une Soufflerie"*. SCS'04, Monastir-Tunisia, Mars 2004.

[13] Hu H. H. and Boau H. H.: *"Feedback Control to Delay or Advance Linear Loss of Stability in Planar Poiseuille Flow"*. Proc. R. Soc. Lond. A, vol. 447, pp. 299-312, 1994.

[14] Lions J. L.: *"Optimal Control of Systems Governed by Partial Differential Equations"*. New York, Springer-Verlag, 1971.

[15] G. P. Liu and S. Daley: *"Optimal Tuning PID Control for Industrial Systems"*. Control Engineering Practice, vol. 9, pp. 1185-1194, 2001.

[16] Y. Mitsukura, T. Yamamato and M. Kaneda: *"A Design of Self Tuning PID Controllers Using A Genetic Algorithm"*. American Control Conference, ACC'99, 1999.

[17] C. Pedret, R. Vilanova and I. Serra: *"A Refinment Procedure for PID Controller Tuning"*. Computers and Chemical Engineering, vol. 26, pp. 903-908, 2002.

[18] F. G. Shinskey: *"PID Dead-time Control of Distributed Processes"*. Control Engineering Practice, vol. 9, pp. 1177-1183, 2001.

[19] G. J. Silva, S. P. Datta and S. P. Bhattacharyya: *"New Results on the Synthesis of PID Controllers"*. IEEE Trans. Automatic Control, vol. 47, no. 2, pp. 241-252, 2002.

[20] Van Der Boom A. J. W. and Van Der Enden A. W. M., The determination of the orders of process and noise dynamics. Automatica, vol. 10, pp. 245-256, 1978.

[21] Wellstead P. E., An instrumental product moment test for model order estimation. Automatica, vol. 14, pp. 89-91, 1978.

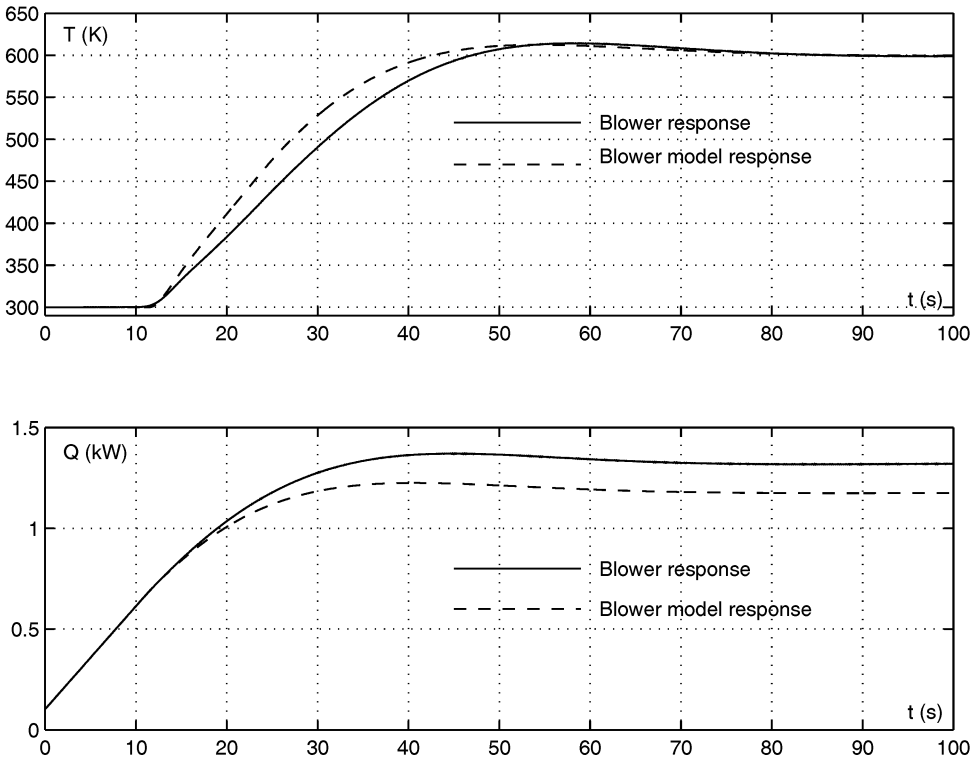


Figure 3: Closed loop responses of the system and the model.

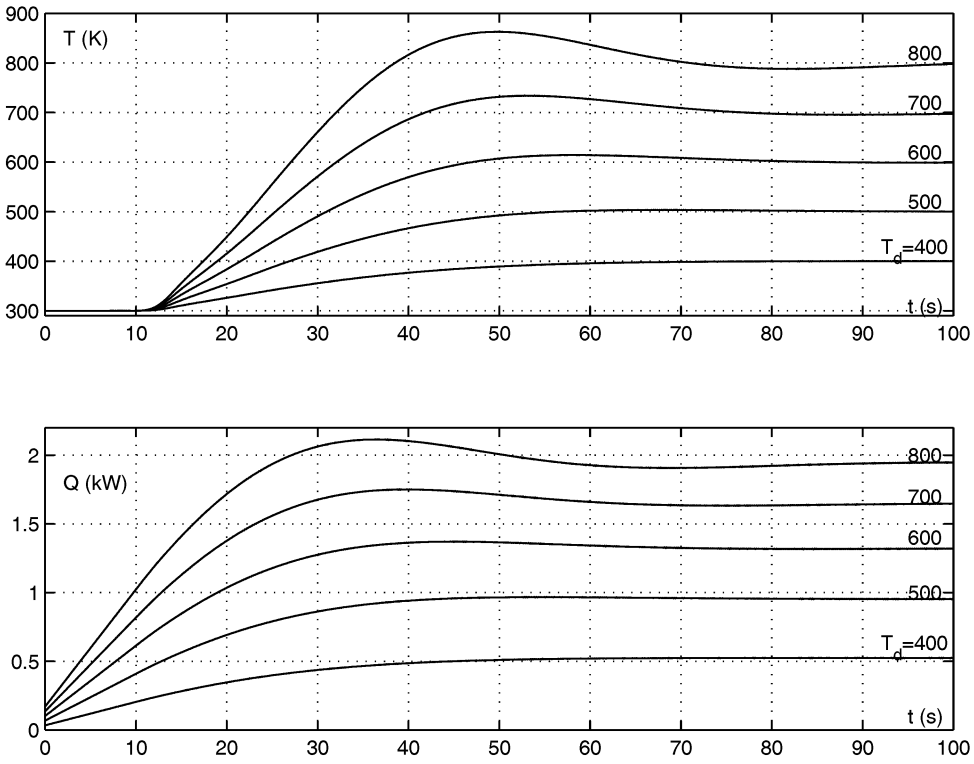


Figure 4: Closed loop responses of the system for different operating points.

On the design of fuzzy supervised PI controller based on gain and phase margin specifications

L.Tounsi-Rekik and M. Chtourou
Research Unit on Intelligent Control, Optimization, design & Optimization of Complex Systems (ICOS)
Department of Electrical Engineering
National School of Engineers of Sfax, BP. W, 3038, Sfax, Tunisia
Tel. (216) 74 675 667, Fax. (216) 74 275 595
Email: rekiklamia@yahoo.fr Mohamed.chtourou@enis.rnu.tn

ABSTRACT

The new proposed PI tuning method using a fuzzy supervisor for a given gain and phase margin specifications is investigated in this paper. The proposed algorithm is a supervised structure where a fuzzy system gives at each sample time the good PI parameters for the regulation of the output of a linear process. Simulation examples are included to confirm the validity and performance of the proposed control algorithm.

Keywords: PID controller, supervision, fuzzy system, gain and phase margin.

1. Introduction

The proportional-plus-integral-plus derivative (PID) controllers have found wide acceptance and applications in the industries for the past few decades [1]. In spite of their simple structures, PID controllers are proven to be sufficient for many practical control problems [1] and hence are particularly appealing to practicing engineers. An abundant amount of research work has been reported in the past on the tuning of PID controllers (Ziegler Nichols step response, Ziegler Nichols ultimate cycling, Cohen-Coon, Internal model control, and error integral criteria).

Gain and phase margin are typical control loop specifications associated with the frequency response technique. They reflect on the performance and stability of the system and are widely used for controller designs.

Added to that, gain margin and phase margin have always served as important measures of robustness [2]. It is also known from classical control theories that phase margin is related to the damping of the system, and therefore also serves as a performance measurement. Controllers designed to satisfy gain margin and phase margin (GPM) criteria are not new approaches. However their solutions are normally obtained numerically or graphically by trial and error use of Bode plots [3]. Such methods are certainly not suitable for use in adaptive control and auto-tuning [4].

Tuning of PID controllers based on gain and phase margin specifications (GPM) was proposed by hang [2]. This method uses linear equations to approximate the arctan function in the gain-phase margin definition, formulas are then derived to design the PI and PID controllers to meet user-specified gain margins and phase margins.

The disadvantage of the (GPM) method is that the transfer function of the controlled process is restricted to the first order plus time delay type.

In this paper we present a new tuning method that uses a fuzzy supervisor based on gain and phase margin specifications, to tune the PI controller parameters more efficiently.

The rest of the paper is organized as follows. In section two we briefly introduce gain and phase margins. Section three proposes the control structure. The synthesis of the fuzzy supervisor is presented in section four. Section five gives the simulation results and discusses the advantage of the proposed approach as compared with other methods. Finally, conclusions are summarized in section 6.

2. Gain and phase margins

Denote the process and controller transfer functions by $G_p(s)$ and $G_c(s)$, and the specified gain margin and phase margin by A_m and ϕ_m, respectively. From the basic definitions of gain margin and phase margin, the following set of equations is obtained:

$$\begin{aligned} &\mathrm{Arg}\left| G_c(jw_p)\, G_p(jw_p) \right| \\ &A_m = \frac{1}{\left| G_c(j\omega_p)\, G_p(j\omega_p) \right|} \end{aligned} \qquad (1)$$

$$\begin{cases} \phi_m = \arg\left[G_c(j\omega_g)G_p(j\omega_g)\right] + \pi \\ \left| G_c(j\omega_g)\, G_p(j\omega_g) \right| = 1 \end{cases} \quad (2)$$

where the gain margin is given by equations (1) and the phase margin by equations (2). The frequency w_p at which the Nyquist curve has an amplitude of $-\pi$ is known in classical control terminology as phase cross over frequency, and the frequency w_g at which the Nyquist curve has an amplitude of 1 as the gain cross over frequency.

It is owing to the complexity of the above equations that an analytical design to achieve specified gain and phase margins is not available to date. Infact in the literature, the achievement of the specified gain and phase margin have been solved numerically or graphically, for example through the use of Bode plots [3].

The PI controller is given by

$$G_c(s) = k_c\left(1 + \frac{1}{sT_i}\right) \quad (3)$$

and the process is given by

$$G_p(s) = \frac{k_p}{1 + s\tau} e^{-sl} \quad (4)$$

From equation (3) and (4), the loop transfert function is

$$G_p(s)\,G_c(s) = \frac{k_c k_p (1 + sT_i)}{sT_i(1 + s\tau)} e^{-sL} \quad (5)$$

Substituting equation (5) into (1)-(2) gives

$$\frac{1}{2}\pi + \arctan(\omega_p T_i) - \arctan(\omega_p \tau_i) - \omega_p L = 0 \quad (6)$$

$$A_m k_c k_p = \omega_p T_i \sqrt{\frac{\omega_p^2\tau^2 + 1}{\omega_p^2 T_i^2 + 1}} \quad (7)$$

$$k_c k_p = \omega_g T_i \sqrt{\frac{\omega_g^2\tau^2 + 1}{\omega_g^2 T_i^2 + 1}} \quad (8)$$

$$\phi_m = \frac{1}{2}\pi + \arctan(\omega_g T_i) - \arctan(\omega_g \tau_i) - \omega_g L$$

Based on the study done in [2], for a given process (k, τ, L) and specifications (A_m, ϕ_m), the parameters of the PI controller are expressed in the following way

$$k_c = \frac{w_p \tau}{A_m k_p} \quad (9)$$

$$T_i = \left(2w_p - \frac{4w_p^2 L}{\pi} + \frac{1}{\tau}\right)^{-1}$$

where

$$w_p = \frac{A_m\phi_m + \frac{1}{2}\pi A_m(A_m - 1)}{(A_m^2 - 1)L} \quad (10)$$

Added to that, the different equations given in [2], can also be used to find the gain margin and phase margins for a given process (k, τ, L) and PI controller parameters (k_c, T_i):

$$A_m = \frac{\pi\tau}{4k_c k_p L}\left(1 + \sqrt{1 - \frac{4L}{\pi T_i} + \frac{4L}{\pi\tau}}\right) \quad (11)$$

$$\phi_m = \frac{1}{2}\pi - \frac{k_c k_p L}{\tau} + \frac{\pi}{4k_c k_p}\left(1 - \frac{\tau}{T_i}\right)$$

The controller synthesis using the method developed in [2], cannot efficiently meet specifications within 20% of overshoot. So to improve the performances, we have adopted another approach using fuzzy systems.

3. The control structure

The following figure shows the adopted control structure

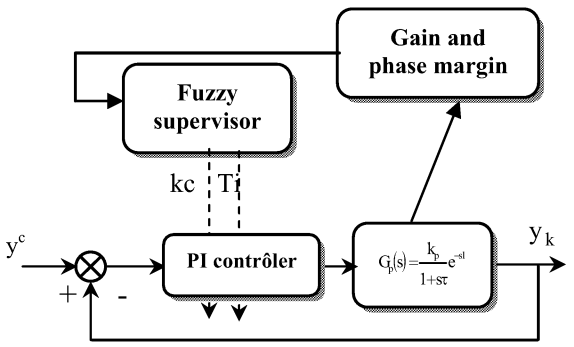


Fig. 1: Control structure

The components of this figure are:

- A classical PI controller giving the control law expressed by the following equation:

$$u(k) = u(k-1) + q_0 e(k) + q_1 e(k-1)$$

where e_k, q_0 and q_1 designate the control error and the PI parameters, respectively that are expressed as:

$$q_0 = -k_c$$
$$q_1 = k_c + T_0 \frac{k_c}{T_i} \quad (12)$$

- A fuzzy supervisor able to provide on-line the PI controller parameters for each pair of gain margin and phase margin (of the considered system).

The fuzzy supervisor is a fuzzy system based on the concepts of fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. In this study, we are interested for a zero order Sugeno fuzzy model used as an approximator [3]. In this study the fuzzy system inputs are the phase margin and the gain margin of the system and two outputs which are the parameters of the PI controller.The i^{th} rule of the inference method is given by :

$$R^i : if \;\; A_m \;\; is \; A_1^i \; and \;\; \phi_m \;\; is \; A_2^i \; then \;\; q_0^i, q_1^i$$

Where i is the rule number. The membership functions have the expression of a parameterized Gaussian function which is specified by two parameters: the center c_i and the width σ_i

$$A_1^i = \exp\left[-\left(\frac{A_m - c_1^i}{\sigma_1^i}\right)^2\right]$$

$$A_2^i = \exp\left[-\left(\frac{\phi_m - c_2^i}{\sigma_2^i}\right)^2\right] \qquad (13)$$

The output of the fuzzy inference system is given by the equation (4).

$$q_l^e = \frac{\sum_{i=1}^{r} \mu_i \, q_l^i}{\sum_{i=1}^{r} \mu_i}, \; l = 0,1 \qquad (14)$$

Where

$$\mu_i = A_1^i A_2^i$$

4. Supervisor synthesis

4.1. Manual synthesis of the supervisor

In the manual approach, many parameters of a fuzzy inference system must be designed by an expert [5]. Hard time can be spent on tuning and constructing fuzzy rules and this becomes more difficult if the required number of rules increases. These motivate researchers to automate the process of fuzzy rule extraction [6] .

4.2 Off line synthesis of the supervisor

The basic idea in the automatic approach is to estimate fuzzy rules through learning from input and output sample data

For training the fuzzy supervisor, we gather 100 points (A_m, ϕ_m), and then we use equations (9) and (10) in order to collect the corresponding values of kc and Ti.

The learning algorithm consists of adjusting the parameter vector containing the pair's c_i, σ_i of each membership function and the terms of the consequent part

The gradient method has been applied to identify the optimal parameter vector, which minimizes the squared sum error committed by the output of the fuzzy inference system $\theta^e = (q_0^e, \; q_1^e)$. For iteration t, the learning algorithm is given by the equation

$$Q(t+1) = Q(t) - \varepsilon \frac{\partial J_p^2(t)}{\partial Q(t)} \qquad (15)$$

$$Q(t) = (\theta^i, d_i, c_i, i = 1...r) \text{ and } \theta^i = (q_0^i, \; q_1^i)$$

represents the consequent parameter to be identified, r is the number of the rules.

The learning algorithm objective is to minimise the following criterion:

$$J_p^1 = \frac{1}{2} \sum_{p=1}^{m} \left\| \theta_p^d - \theta_p^e \right\|^2 \qquad (16)$$

ε, m are the learning rate and the example number, respectively. θ_p represent the values collected from the data base.

The second term of (8) represents the adjustment of the parameter vector after the presentation of the example p to the fuzzy inference system. It is given by (10), (11) and (12).

$$\frac{\partial J_p^1}{\partial c_l^i(t)} = \sum_{l=0}^{2} \frac{\partial J_p^1}{\partial q_l^e} \frac{\partial q_l^e}{\partial \mu_i} \frac{\partial \mu_i}{\partial c_l^i(t)} \qquad (17)$$

$$\frac{\partial J_p^1}{\partial \sigma_l^i(t)} = \sum_{l=0}^{2} \frac{\partial J_p^1}{\partial q_l^e} \frac{\partial q_l^e}{\partial \mu_i} \frac{\partial \mu_i}{\partial \sigma_l^i(t)} \qquad (18)$$

The updating of the consequent parameters of the fuzzy rules is given by the following equation:

$$q_l^i(t+1) = q_l^i(t) - \varepsilon \left(\frac{\mu_i}{\sum_{i=1}^{r} \mu_i} \left(q_l^e - q_l^d \right) \right) \qquad (19)$$

Where l=0,1.

5. Simulation studies

In this section, we give a specific performance comparison with Ziegler Nichols method

The process to be controlled is a first order with time delay and it is given as: [2]

$$G_p(s) = \frac{1}{1+s} e^{-0.1s} \qquad (20)$$

In our work we have adopted the discrete model of the considered process. So the output of the system is given by the following expression

$$y(k) = -a_1 y(k-1) + b_1 u(k-2)$$

where

$$a_1 = -\exp\left(\frac{-T_0}{\tau}\right) \text{ and } b_1 = k_p\left[1 + a_1\right]$$

T_0 =0.05 s is the sampling time.

5.1. Manual synthesis of the supervisor

Simulations results are compared with a conventional PI controller adjust by the method of Ziegler-Nichols.
Figures 2-a and 2-b show respectively the evolution of the uncompensated plant's output and the response of the plant obtained by the control structure when the supervisor has been identified manually. The result of using Zigler Nichols tuning formula to adjust the parameter of the PI controller is given in figure 2-c.

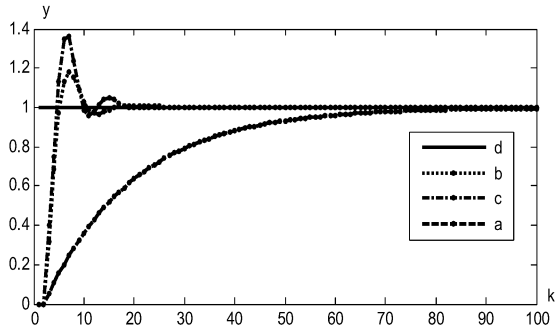


Fig. 2: Evolution of the closed loop output (2-a: uncompensated , 2-b: supervised PI , 2-c: Conventional PI)

It can be seen that the manual synthesis of the supervisor performs much better then the Ziegler Nichols design. In fact the adopted approach has such a shorter rise time and lower overshoot than the Ziegler Nichols tuning formula.

In order to test the robustness of the fuzzy supervisor synthesis manually and to show the validity of the tuning algorithm, we have varied the plant gain kp at the instant k=50 . Simulation results are shown in figure 3.

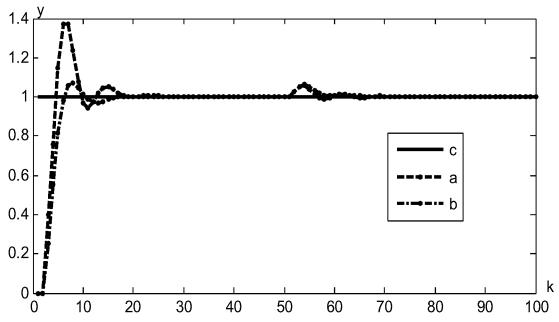


Fig. 3: Evolution of the closed loop output (3-a: conventional Pi , 3-b: supervised PI , 3-c: desired output)

Comparing to the Ziegler Nichols method, it' s clear that the proposed one can improve the system performance with faster response and faster convergence, added to that the system may be more robust to parameter variation.

5.2. Off-line synthesis of the supervisor

In order to improve the performances achieved through the manual synthesis, we have considered the offline synthesis. The result of this approach is plotted in figure 4.

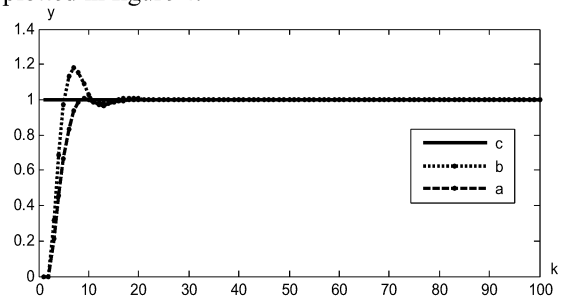


Fig. 4: Evolution of the closed loop output (4-a: off line synthesis of the supervisor, 4-b: manual synthesis of the supervisor, 4-c; desired output)

The graph shows that the off line synthesis of the supervisor gives batter performance than the manual synthesis. In fact, we note that the overshoot of the system, is suppressed automatically.

In order to test the robustness of the fuzzy supervisor synthesis off line, we have varied the plant gain kp at the instant k=50. Simulation results are shown in figure 5.

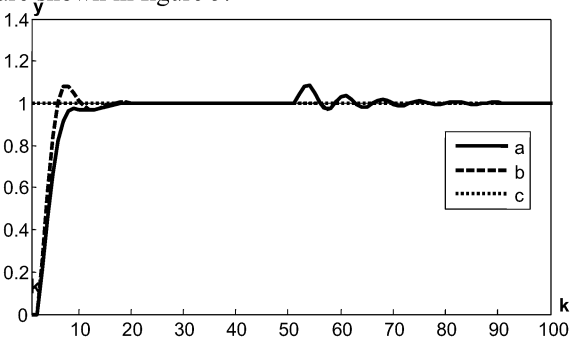


Fig. 5: Evolution of the closed loop output (5-a: off line Synthesis of the supervisor, 5-b: manual synthesis of the supervisor, 5-c; desired output)

We remark that, the tuning algorithm using fuzzy supervisor synthesis off line is so robust to a parametric variation.

6. Conclusion

This paper has investigated fuzzy supervised PI controller based on gain and phase margins. There are two advantages of using the proposed approach, first, the automatic synthesis of the supervisor tunes PI controller parameters for different gain and phase margin specifications so that neither numerical methods nor graphical methods need be used. Second, the new tuning method gives better performance than the conventional one.

References

[1] K. J. Astrom and T. Hagglund *" Automatic Tuning of PID Controllers",* Automatica, 2nd edn. Instrument Society of America. NC, USA 1995.

[2] Weng Khuen Ho & Chang Chieh Hang & Lisheng S. Cao *"Tuning of PID Controllers Based on Gain and Phase Margin Specifications",* Automatica, Vol. 31, N° 3, pp. 497-502, 1995.

[3] Ho-Wang Fung & Qing-Guo Wang & Tong-Heng Lee *"PI Tuning in Terms of Gain and Phase Margins",* Automatica, Vol. 34, N° 9, pp. 1145-1149, 1998.

[4] W. K. Ho & K. W. Lim & Wen Xu *" Tuning of PID controllers based on gain and phase margin specifications using fuzzy neural network",* Fuzzy sets and systems, Vol. 101, N°3, pp. 21-30, 1999.

[5] T. Takagi, M. Sugeno "*Fuzzy identification of systems and its applications to modelling and control*", IEEE Trans. on systems Man and Cybernetics, vol. SMC-15, n°1, pp. 116-132, 1985.

through Linearization", IEE Proc., control applications, 1995, pp. 315-322.

.**[6]** L-X.Wang and J.M.Mendel, "*Generating fuzzy rules by learning examples*", IEEE Trans. Systems Man and Cyber. SMC-22, pp. 1414-1427, 1992.

PID NEURAL TUNING FOR INTEGRATING PROCESSES

R. Chibani and M. Chtourou
Research unit on Intelligent Control, Optimization, design & Optimization of Complex systems (ICOS)
Department of Electrical Engineering
National School of Engineers of Sfax, BP. W, 3038, Sfax, Tunisia.
Tel. (216) 74 274 088, Fax. (216) 74 275 595
Email: chi_radhia@yahoo.fr, *Mohamed.chtourou@enis.rnu.tn*

Abstract

This paper presents a new PID controller method for integrating processes with time delay and time constant. This method is based on the effective combinations of classical PID controllers and emerging neural networks as intelligent technologies. First, the neural supervisor is designed off line using the technique of Wang and al. Then, the obtained supervisor is introduced in the loop to be used on line to adjust the PID parameters. Simulation examples are given to illustrate the proposed method.

Keywords: PID controller; auto-tuning; neural networks, integrating process; time delay; robustness.

1. Introduction

Despite a lot of research and the huge number of different solutions proposed, most industrial control systems are still based on conventional PID controllers. Different sources estimate the share taken by PID controllers at between 90 and 99% [1]. Some of the reasons for this situation are given as follows. Firstly, PID controllers are robust and simple to design. Moreover, there exists a clear relationship between PID and systems response parameters. As a PID controller has only three parameters, plant operators have a deep knowledge about the influence of these parameters and the specified response characteristics on each other. Secondly, many PID techniques have been elaborated [2] such as; Ziegler-Nichols methods based on the open-loop step response and on the frequency response. Finally, because of its flexibility, PID control could benefit from the advances in technology. Most of the classical industrial controllers have been provided with special procedures to automate the adjustment of their parameters (tuning and self-tuning) [3]-[6].

However, PID controller cannot provide a general solution for all control problems. The processes involved are in general complex and time-varying, with delay and non-linearity, and often with poorly defined dynamics. In addition, it is very difficult to control integrating processes with time delay. Therefore, many tuning methods have been proposed.

Zhang and al. [7] have proposed a PID controller design method for integrating processes with time delay and time constant. Firstly, they investigate the conditions that guarantee the stability of the closed-loop system. Then, an optimization method using H_∞ performance criterion is developed for deriving the controller.

Chidambaram and al. [8] have developed a simple method for the PI, PD and PID controller settings for an integrating plus dead-time transfer function model.

Chien and al. [9] have developed a simple modified Smith predictor controller design for integrating processes with long dead-time.

However, they usually show a large overshoot for a step set point change and poor disturbance attenuation due to the integrator of the process [10]. In this paper, a method of designing PID controller has been developed for the control integrating processes with time-delay and time constant. This approach is based on the effective combinations of classical PID controller and emerging neural networks.

This paper is organized in the following manner. In Section 2, the proposed method is presented. Simulation examples will follow in Section 3 to illustrate the use of the method. Conclusions will be drawn in Section 4.

2. Proposed method

Let us consider the integrator plus dead-time and time constant and controller transfer functions given by:

$$G_p(s) = \frac{K}{s(Ts+1)} e^{-Ls} \quad (1)$$

$$C_c(s) = k_p + \frac{k_i}{s} + k_d s \quad (2)$$

where s is the Laplace operator. K, T and L denote the static gain, the time constant and the time delay respectively.

This method is based on the use of a neural supervisor to adjust on-line the PID parameters so that the process response matches with the desired output.

The supervisor has the sensitivity, the static gain, time constant and the time delay of the process as inputs and the three PID parameters as outputs.

Before being used, the supervisor has been trained off-line with a data set formed by the couple $\{(M_s, L, K, T), (k_p, k_i, k_d)\}$. The collection of the data base requires the determination of the PID parameters, for each process $\{M_s, L, K, T\}$. Here, we are based on the Wang and al. formulas to tune the PID controller for each integrating process with time delay to meet sensitivity specification [10].

The procedure of this technique is represented by the following scheme:

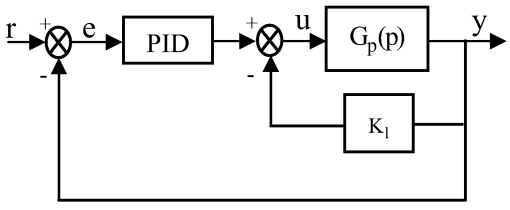


Fig.1: Controller black diagram for integrating process

We remark here the use of the P controller (K_1) in the inner feedback loop which permits to change the integrating process to a stable process.

Sun and Lee [10] derived the P controller gain for the optimum disturbance rejection, which is:

$$K_1 = \frac{0.2}{KL} \quad (3)$$

Then, the PID controller of the outer loop is designed based on sensitivity specification defined as:

$$M_s = \max_{0 \le w \le \infty} \left| \frac{1}{1 + G_p(jw) C_c(jw)} \right| \quad (4)$$

where the loop function is $G_1(jw) = C_c(jw) G_p(jw)$. The quantity M_s is the shortest distance from the Nyquist curve of the loop function $G_1(jw)$ to the critical point $(-1, j0)$. This means that the Nyquist curve of the loop transfer function $G_1(jw)$ is tangent to a circle around the critical point with radius $1/M_s$. Therefore, using this condition, the PID parameters are determinate by the following expressions:

$$k_p = \frac{1}{KL}\left(1.1608 - \frac{1.2064}{M_s}\right) \quad (5)$$

$$k_i = \frac{1}{KL^2}\left(0.29 - \frac{0.3016}{M_s}\right) \quad (6)$$

$$k_d = \frac{T + 0.1L}{KL}\left(1.450 - \frac{1.508}{M_s}\right) \quad (7)$$

The data base obtained has been used for the synthesis of the neural supervisor. The neural network training is carried out by using the algorithm of backpropagation. The training of this network consists in seeking the connection values to minimize a function which is the sum of the squares of the errors made on the entire data base:

$$E^p(w) = \sum_{i=1}^{DB} \frac{1}{2} \left\| \theta_i^{RN} - \theta_i^d \right\|^2 \quad (8)$$

where: θ^d denote the desired values contained in the data set and θ_i^{RN} indicates the PID parameters estimated by the neural supervisor.

After the training phase, the supervisor is introduced in the closed loop with the conventional PID and the system.

3. Simulation examples

In the following, we will give the comparisons between the proposed PID tuning method and other methods such as Wang's method [10] and Zhang's method [7].

Consider the following an integrating process with a time delay:

$$G_p(s) = \frac{1}{s(s+1)} e^{-2s} \qquad (9)$$

Fig. 2 shows the closed loop process output responses for these three designs with a unit-step setpoint change and distrubance occuring at t = 50s (proposed method, solid line; Wang's method, dotted line; Zhang's method, dashed line). The proposed method shows better control performance both for setpoint and load disturbance.

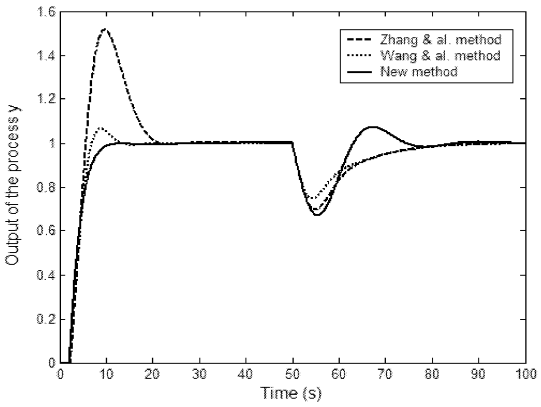


Fig.2: Process response with other tuning methods

In order to compare the proposed method to Wang's method and Zhang's method, a variable time delay of the integrating process has been considered. The process considered has the following expression:

$$G_p(s) = \frac{1}{s(s+1)} e^{-s} \qquad (10)$$

Using the neural supervisor designed previously to tune the PID controller for the integrating process with time delay and time constant; the results are shown in Figure 3. This figure presents the closed loop process output responses for these desgins with a disturbance occuring at t = 50s.

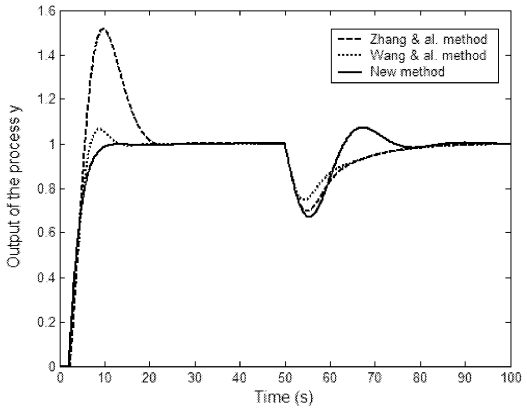


Fig.3: Process response with other tuning methods

The propose method shows the better control performance both for the setpoint change and load disturbance (proposed method, solid line; Wang's method, dotted line; Zhang's method, dashed line).

For purpose of comparison, we have used the same neural supervisor for the control of the integrating process with a variable time delay given below:

$$G_p(s) = \frac{1}{s(s+1)} e^{-0.8s} \qquad (11)$$

The control performance of the proposed method is compared with the result of Wang's method and Zhang's method, as shown in Fig. 4, where a load disturbance is added at 50s (proposed method, solid line; Wang's method, dotted line; Zhang's method, dashed line).

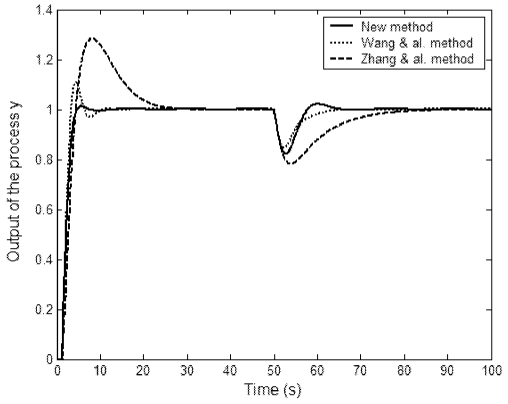


Fig.4: Process response with other tuning methods

4. Conclusions

In this paper, the control problem of integrating processes involving time delay and time constant is discussed. A new PID controller method for integrating processes with time delay and time constant is proposed.This method is based on the effective combinations of classical PID controllers

and emerging neural networks as intelligent technologies. First, the neural supervisor is designed off line using the technique of Wang and al. Then, the obtained supervisor is introduced in the loop to be used on line to adjust the PID parameters. According to simulation results, the proposed method gives better performance in the setpoint and load distrubance than the two other designs methods.

References

[1] K. J. Silva, A. Datta and S. P. Bhattacharyya"*New Results on the Synthesis of PID Controllers",* IEEE Trans. On Automatic Control, Vol. 47, N°2, pp. 241-252, February 2002.

[2] G. J. Silva, A. Datta and S. P. Bhattacharyya"*New Results on the Synthesis of PID Controllers",* IEEE Trans. On Automatic Control, Vol. 47, N°2, pp. 241-252, February 2002.

[2] K. J. Aström & T. Hägglund "*PID controllers: Theory, Design and Tuning",* Instrument Society of America, Research Triangle Park, North Carolina, 1995.

[3] P. Cominos & N. Nuro *"PID Controllers: Recent Tuning Methods and Design to Specification",* IEE Pro.-Control Appl., Vol. 149, No. 1, January 2002.

[4] T. Hägglund an K. J. Aström *"Revising the Ziegler-Nichols Tuning Rules for PI Control"*, Asian Journal of Control, Vol.4, N°4, pp. 364-380, December 2002.

[5] B. Kristiansson and B. Lennartson *"Robust Optimal Tuning of PI and PID Contollers"*. IEE Proc. Control Theory Appl., Vol.149, N°1, pp. 17-25, January 2002.

[6] G. K. I. Mann, B.-G. Hu and R. G. Gossine *"Time-domain based Design and Analysis of New PID Tuning Rules"*. IEE Proc. Control Theory Appl., Vol.148, N°3, pp.251-262, May 2001.

[7] W. Zhang, X. Xu and Y. Sun *"Quantitative Performance Design for Integrating Processes with Time delay",* Automatica, Vol. 35, pp. 719-723, 1999.

[8] I. L. Chien, S. C. Peng and J. H. Liu *"Simple control method for Integrating processes with Long Deadtime",* Journal of Process Control, Vol. 12, pp. 391-404, 2002.

[9] M. Chidambaram and R. P. Sree*"A Simple Method of Tuning PID Controllers for Integrator/dead-time processes", Computers and Chemical Engineering*, Vol. 27, pp. 211-215, 2003.

[10] Y. G. Wang and W. J. Cai *"PID Tuning for Integrating Processes with Sensitivity Specification",* Proc. of the 40th IEEE Conference on Decision and Control Orlando, Florida USA, December 2001.

PHONE NETWORK SIMULATION

SERVICE-AWARE ROUTING PROTOCOLS (SARP): EXPLOITING MOBILITY WITH LOOSE DELAY CONSTRAINTS IN AD HOC NETWORKS

M. Alchaita
M. Al-Akaidi
J. Ivins
School of Engineering & Technology
De Montfort University
Leicester, UK
E-mail: mma@dmu.ac.uk

KEYWORDS
Mobile Ad-hoc Networks, Routing Protocols, Heading Directional Angle, Service-aware, Mobility

ABSTRACT

In this paper, we present two routing algorithms for mobile ad-hoc networks, which exploit the mobility of nodes to reduce flooding and overhead in the network and to avoid the mutual interference of concurrent transmissions between nodes. These algorithms give solutions to some of the problems encountered in ad hoc networks, such as handling the frequent changes in the network topology due to mobility; and overhead traffic controls. These algorithms designed for applications that are asynchronous in nature with end-to-end tolerant delays. Based on their Heading Direction Angle (HDA), only a small number of the nodes are selected to forward a message from the source to the destination. Under this assumption, these algorithms reduced the overhead and flooding in the network when nodes are more mobile rather than stationary.

INTRODUCTION

A mobile ad hoc network (MANET) is a collection of mobile nodes where communication is established in the absence of any fixed infrastructure. The only possible direct communication is between neighboring nodes; therefore communication between remote nodes is based on multi hop. Each mobile host acts as a host and a router, relaying information from one neighbour to others. Host mobility can cause frequent unpredictable topology changes. Therefore, the task of finding and maintaining routes in MANET is nontrivial. Many protocols have been proposed for mobile ad hoc networks, with the goal of achieving efficient routing (Basagni et al. 2001; Dube et al. 1997; Toh 1999; Elizabeth M. Royer 1999; Charles E. Perkins et al. 2003; Blazevic et al. 2005).

Communication between arbitrary end-points in an ad-hoc network requires routing over multiple-hop wireless paths. The main difficulty arises because without a fixed infrastructure these paths consist of wireless links whose end-points are likely to be moving independently of one another. Consequently, node mobility causes the frequent failure and activation of links, leading to increased network congestion while the network's routing algorithm reacts to the topology changes. Unlike fixed infrastructure networks where link failures are comparatively rare events, the rate of link failure due to node mobility is the primary obstacle to routing in ad-hoc networks.

Several mobility models have been described that represent mobile nodes whose movements are independent of each other (e.g., the Random Walk Mobility Model and the Random Waypoint Mobility Model) and several mobility models that represent mobile nodes whose movements are dependent on each other (e.g., group mobility models)(Camp 2002; Musolesi et al. 2004; Guolong Lin 2004). A routing protocol that uses geographical information (GPS information) (Samarth 2002; Young-Bae Ko 2000; Kumar Banka 2002) may not work well in certain situations where such information is not available.

Mobility in ad hoc networks is an obligation that cannot be obviated; on the other hand, the mobility causes the expiration of the information used by traditional routing protocols in establishing a route. As mobility is inevitable, we need to exploit the positive aspect of this phenomenon as much as possible. One of the positive aspects in mobility is the direction of the node. (Gupta and Kumar 2000) present a model for studying the capacity of *fixed* ad hoc networks, where all nodes are randomly located in the network area but are stationary. (Grossglauser and Tse 2002) proposed a model of an ad hoc network that exploits mobility of nodes by using a form of multiuser diversity via packet relaying .In this paper; we suggest two approaches that exploit the mobility of nodes to hand the packets off to the final destination using the heading direction information for mobile hosts.

ALGORITHMS DESCRIPTION

The purpose of our approaches is to improve the service-aware routing protocols in Mobile Ad hoc Networks. In our work, we focus on applications that are asynchronous in nature with accepted and predetermined tolerance of end-to-end delays, hence, increase network capacity by exploiting nodes mobility. On such long time-scale applications, even more network performance gain can be obtained because the changes of network topology will not

affect packets transmissions over time due to user mobility; instead, mobility will help packets to be delivered to the destination. Examples of such applications that accept some delay include electronic mail, advertising events, and certain types of event notification. Our strategy is for each source node to send its data packets to different selected nodes depending on the information provided to the source node (e.g., destination location), and on the situation of the network (stationary, mobile most of the time,…). The time that the packet should arrive to the destination *PT* is included in data packet (the period that the packet is allowed to be delayed from the destination). Selecting these nodes will be according to heading direction angle and location of destination. These nodes then serve as mobile relays and whenever they get close to the final destination, they hand the packets off to the final destination. Every interval time the node that carry the packet checks *PT*, if *PT* is about to expire or the node changes its direction, the node will hand the packet off to another selected node that is near to the destination (if nodes locations are known) and moves toward the destination.

Each one of proposed algorithms could be divided into two parts: the mobility part and relaying messages from the source node *S* to the final destination *D* is the second part (we have denoted this part as service-aware part). The former concerns the mobility of nodes and the classification of neighboring nodes, where each node classifies its neighbors into eight groups depending on the heading direction angles of these nodes, and both proposed algorithms share this part. The latter has the main task of transmitting packets from the source to the destination and choosing the way of forwarding packets to other nodes during packet transmission task; this part distinguishes between the two proposed approaches. Next sections discuss these parts in details.

Mobility and classification of nodes

This part of algorithm is shared by both proposed algorithms, and it makes use of direction information of nodes in the network to reduce routing overhead and elongate the life time of links between communicated nodes. Heading Direction angle (HDA) may be provided by a digital compass. It has been assumed that each mobile node in the network is equipped with a digital compass with magneto resistive (MR) technology (Solid State Electronics Center, Home page), which delivers the heading direction angle of the mobile device hosting it. It has also been assumed that each node will exchange heading direction information with its neighbors in a periodic manner. The information received from another node (neighbor) will be stored in one of the eight zones ranges in the cache regardless of the actual position of that neighbor.
Each mobile node in the network will classify its neighbor's nodes into eight different zone areas (d1… d8) according to the heading direction of those neighbors, as shown in Figure 1; theoretically, the nodes are categorized within the eight zones ranges, regardless of their actual relevant positions to the node.

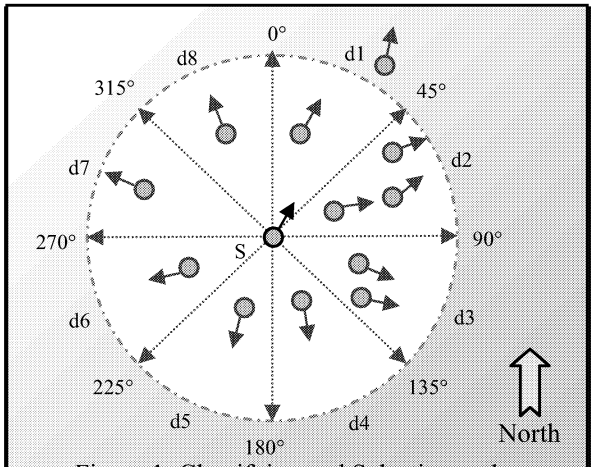


Figure 1: Classifying and Selecting nodes according to HDA.

Service-Aware Schemes

Scheme1:

In this scheme, we assume that the location of destination *D* and other nodes in the networks are not known, and each node is aware about the heading direction angle of its neighbors, Destination could be a node or a determined region. We also assume that each data packet sent from higher layers to network layer includes the time *PT* that the packet should arrive to the destination (in another word, *PT* is the tolerance delay allowed for packet transmission). Mobile nodes can keep copies of packets they receive and forward these packets to other nodes in three cases:

- When the node changes its heading direction angle from one direction to another.
- If *PT* is about to expire.
- If *D* becomes a neighbour to the mobile node that keeps the packets, the node will hand packets off to *D*.

Each packet to be delivered will almost certainly have many copies saved in the network nodes at a given instant.
When *S* has data packets to send to *D*; *S* will select one neighbor from each of the eight heading direction zones that are depicted in figure 1. If a neighbor is found in the intended zone, copy of data packets is sent to that neighbor, otherwise, two neighbors will be selected from the following zone to maintain each one a copy of data packets. The first neighbor will maintain a copy of data packets with its zone information (Zone Number), whilst the second neighbor will maintain a copy of data packets with the previous zone information (previous zone number), and on the time-scale whenever this neighbor has a neighbor belong to previous zone, it hands the copy of data packets off to that neighbor. And the same rules will be followed if neighbors from two adjacent zones are not found, three neighbors will be selected from the third zone, first node from these three nodes will maintain a copy of data packets and its zone information, and the two other nodes will maintain information about the previous zones that have no neighbors.

Now, after *S* hands copies of data packets off to its neighbors, each neighbor *Nb* will keep a copy in its cache, and at every interval time, *Nb* checks the *PT* of data packet carried in its cache, if *PT* is about to expire *Nb* will select a neighbour from the same zone that *Nb* fall into, and follow the same rules followed by *S* in case the neighbour is not found. Then *Nb* will hand the packet off to that neighbour. When the node *MN* approaching the destination *D* that *MN* has packet for, and becomes a neighbour to *D*, *MN* hands the packets off to *D*. As we can see that each node can keep data packets for more than one destination in the same time. The general flowchart for scheme1 is shown in figure 2.

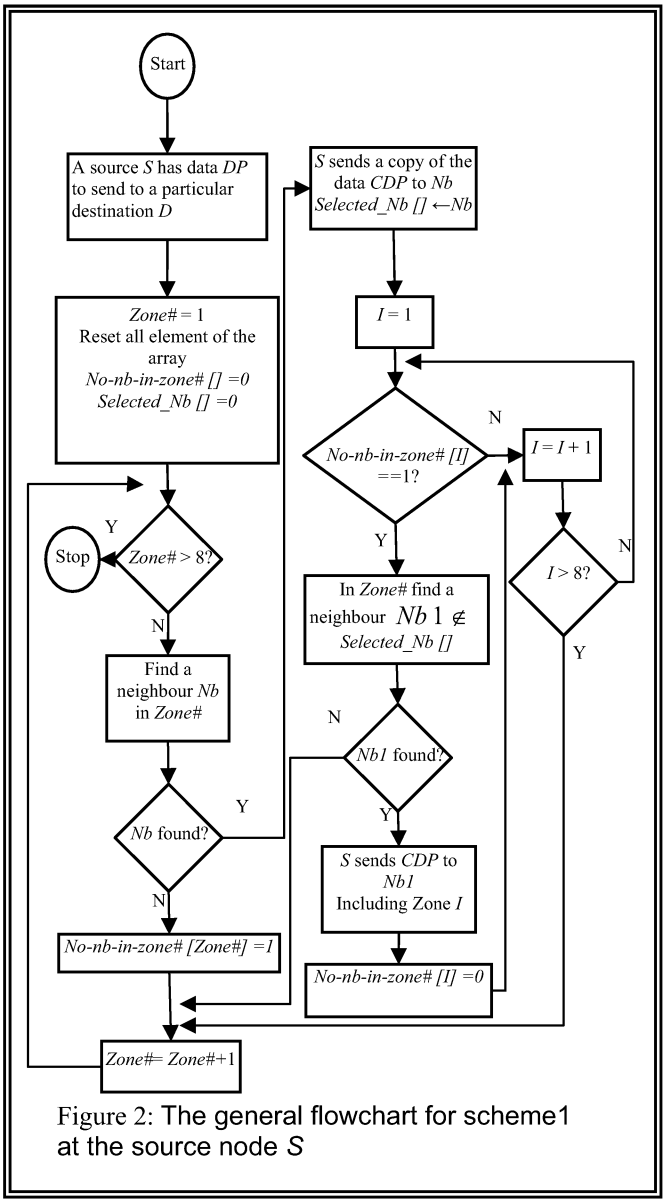


Figure 2: The general flowchart for scheme1 at the source node *S*

Scheme2:

In this scheme, we assume that the location of the destination *D* is known by the source node *S* (D could be a mobile node or a determined region contains at least one node), and each node knows the geographical information of its neighbors. The geographical information could be obtained by the Global Positioning System (GPS). In addition, each node is aware about the heading direction angle of its neighbors, and each data packet sent from higher layers to network layer includes the tolerance delay time *PT* allowed for packet transmission. Mobile nodes can keep copies of packets they receive and forward these packets to other nodes in the following cases:

- When the node that carries a copy of data packet for particular destination *D* moves far from that destination.
- When the node changes its heading direction angle from one direction to another and the new direction leads the node to be far from *D*.
- When *PT* is near to be expire.
- If *D* becomes a neighbour to the mobile node that keeps the packets, the node will hand packets off to *D*.

Transmissions are constrained to nodes with heading direction angles toward *D* and closer than the distance between *S* and *D*. These nodes then serve as mobile relays. Thus, the network is not necessary to be connected at any given instant and changes in network topology are not essential for packet delivery. The basic idea of our work depends on using many different relay nodes, where at least one relay will hopefully get close to the destination. Although the basic communication problem in ad hoc networks is point-to-point, the approaches we proposed effectively solve this problem by relying on only the nodes that keep a copy of packets.

When *S* has data packets to send to *D*; *S* will select some neighbors (two or more neighbors) from neighbors available in the heading direction zone that have heading direction toward the destination location. The selected number of neighbors depends on the network density and the average mobility speed. Figure 3 depicts the way of selecting the route to destination.

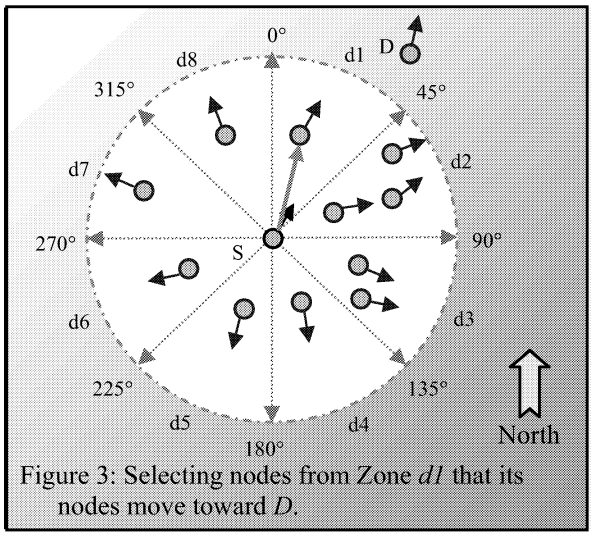


Figure 3: Selecting nodes from Zone *d1* that its nodes move toward *D*.

If neighbors are found in the intended zone, copies of data packets *CDP* are sent to those neighbors, otherwise, another neighbors will be selected from the following zone(s) to maintain (each one of them) a copy of data packets, with

including the location information of *D* in the packet. On the time-scale, and after *S* hands copies of data packets off to its neighbors, each neighbor *Nb* will keep a copy in its cache, and at every interval time, *Nb* checks the *PT* of data packet carried in its cache, and check the distance between itself and *D*, and:

- If *PT* is about to expire *Nb* will select a neighbour that is moving toward *D* or closer to *D*, and follow the same rules followed by *S* in case the neighbour is not found in a zone.
- If the node notices that is move away from *D* by the time, it will select a neighbour that is moving toward *D* or closer to *D*.

Then *Nb* hands the packet off to that neighbour. When the node *MN* approaching the destination *D* that *MN* has packet for, and becomes a neighbour to *D*, *MN* hands the packets off to *D*. The general flowchart for scheme2 is shown in figure 4.

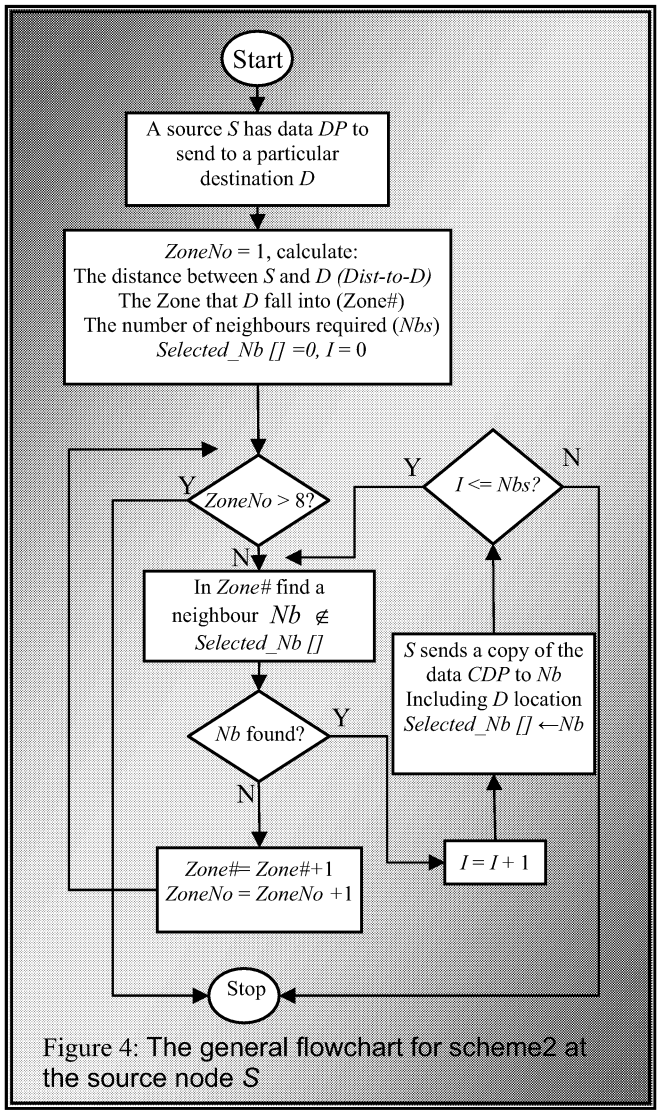


Figure 4: The general flowchart for scheme2 at the source node *S*

RESULTS AND DISCUSSION

The network simulator ns-2 was used to perform extensive simulations and to evaluate our Network. A total of 50 nodes were simulated for 500 seconds over a network space of 1000x1000m. The network traffic was modelled as 10 CBR sources. Movement patterns were generated based on a random waypoint model for each value of pause time and node speed. As expected our algorithms has reduced the overhead traffic control with different pause time and different node speed, due to selecting a small number of nodes to forward packets in the network, and because selecting the neighbor that fulfill the conditions of heading direction angles. Results will be presented in the conference time.

CONCLUSION

In this paper, we have presented two service-aware routing algorithms for mobile ad-hoc networks. These algorithms are designed for applications with loose delay such as email service and text messages. The strategy depends on determine the time allowed for packet to be delayed from the destination, and exploit the mobility of nodes to reduce flooding and overhead in the network and to avoid the mutual interference of concurrent transmissions between nodes. In such applications with tolerance delay, these algorithms handle the frequent changes in the network topology due to mobility; and overhead traffic controls.

REFERENCES

Basagni S., Turgut D. and Das S. K. 2001, "Mobility-adaptive protocols for managing large ad hoc networks", *Proceedings of IEEE International Conference on Communications (ICC)*, Jun, 1539-1543.

Blazevic L.; Le Boudec J.-Y.; Giordano S. 2005, "A location-based routing method for mobile ad hoc networks"; *Mobile Computing, IEEE Transactions,* Volume 4, Issue 2, March-April Page(s):97 - 110

Camp T., Boleng J., and Davies V. 2002, "A survey of mobility models for ad hoc network research", *Wireless Communications and Mobile Computing (WCMC)*: Special issue on Mobile Ad Hoc Networking,483-502.

Charles E. Perkins, Elizabeth M. Belding-Royer, Samir R. Das 2003, "Ad hoc On-Demand Distance Vector (AODV) Routing", *Internet-draft, in: Mobile Ad-hoc Network (MANET) Working Group*, IETF, 17 February.

Dube R., Rais C.D., Wang K-Y. and Tripathi S.K. 1997, "Signal Stability-Based Adaptive Routing (SSA) for Ad-Hoc Mobile Networks", *IEEE Personal Communications*, February, 36-45.

Elizabeth M. Royer 1999, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks", *IEEE Personal Communication,* April.

Guolong Lin, Guevara Noubir, and Rajmohan Rajamaran 2004, "Mobility models for ad-hoc network simulation", *In Proceedings of IEEE Infocom.*

Kumar Banka, R. Guoliang Xue 2002, " Angle routing protocol: location aided routing for mobile ad-hoc networks using dynamic angle selection.", *IEEE Milcom'2002 Conference*, vol.1, 501- 506.

Musolesi M., Hailes S., C. Mascolo 2004,"An ad hoc mobility model founded on social network theory", *In Proceedings of the 7th ACM/IEEE International Symposium on Modeling,*

Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2004), October.

Solid State Electronics Center, home page, http://www.ssec.honeywell.com/magnetic/.

Samarth H., and Klara Nahrstedt 2002, "Predictive Location-Based QoS Routing in Mobile Ad Hoc Networks", *In Proceedings of IEEE International Conference on Communications*.

Toh C-K. 1999, "Associativity-Based Routing for Ad-Hoc Networks", *Wireless Personal Communications Journal, Special Issue on Ad-Hoc Networks*, Vol. 17, No. 8, August, 1466-1487.

Young-Bae Ko and Nitin H. Vaidya 2000, "Location-Aided Routing (LAR) in mobile ad hoc networks", *Wireless Networks 6*, 307–321.

Grossglauser M. and Tse 2002, "Mobility increases the capacity of ad hoc wireless networks" *Networking, IEEE/ACM Transactions on* Volume 10, Issue 4, Aug. 2002 Page(s):477 - 486 .

Gupta P. and Kumar P. R. 2000, "The capacity of wireless networks", *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.

Effects of Handoffs on the Performance of Cellular Systems Using Simulation

Nada Al-Khatib
Communication & Electronics Eng. Dept.
Philadelphia University, Jordan
E-mail: nadakhtb@yahoo.com

Jamal S. Rahhal
Electrical Engineering Dept.
University of Jordan
E-mail: rahhal@ju.edu.jo

KEYWORDS
Communications, Engineering, Simulation, Performance analysis, Global, probabilistic

ABSTRACT

One important performance factor in Cellular systems is the Handoffs. As the handoffs increases, the overhead on the system increases and simultaneously for a fraction of a second two channels will be reserved for one user and this will reduce the overall systems capacity. In this paper, the time of holding two channels during handoff execution in synchronous and asynchronous handoff was found and the effects of different aspects of handoffs on systems capacity are calculated for various situations by using simulation. It is shown that in small cell up to 16 % of capacity reduction could result due to high handoffs rates. Therefore, reducing probability of handoffs will enhance the system's capacity.

INTRODUCTION

GSM (Global System for Mobile Communications) is the most common Second Generation mobile telephony standard. GSM has been chosen as one of the standards from which 3G mobile telecommunications will evolve. Nowadays, the GSM core network evolved with GPRS packet data functionality. GPRS (General Packet Radio Service) is an enhancement of core GSM networks, that allows the rapid transfer of data bundled into 'packets', separate from voice or data call circuits. It is a 2.5G stepping stone to 3G networks and services, GPRS involves overlaying a packet based air interface on the existing circuit switched GSM network. This gives the user an option to use a packet-based data service. To supplement circuit switched network architecture with packet switching needs only to add a couple of new infrastructure nodes and making a software upgrade to some existing network elements.
In GSM, GSM/GPRS or UMTS (3rd generation mobile system) the mobility management is the key factor to achieve wireless mobile communication which it involves two operations: location management and handoffs management (Chung-Ming Huang 2004).
Handoffs methods are required in wireless mobile communication to facilate providing continuous service for population of moving Mobile station which may move between cells, between frequencies and between systems. Properly designed handoff procedure is essential in keeping as low as possible the signaling and switching load to the network. One way of finding how much capacity reduction of cellular communication system could result due to handoffs rate is using simulation.
A simulation generally refers to a computerized version of the simplified representation of actual system that runs over time to generate some output to predict the behavior of the system. For example, simulation of a typical cellular system then changing one of the different parameters in the cellular system to show how the number of handoffs is affected when the parameter change in the real system and study (by analyzing the output results of the simulation) the capacity reduction due to handoffs rate.

WASTED TIME DUE TO HANDOFFS

Handoffs require holding of two channels for certain time period as mentioned previously. Figure (1) shows a handoff process between channels of the same BSC, only after activation of TCH in new cell and before releasing the old channel.

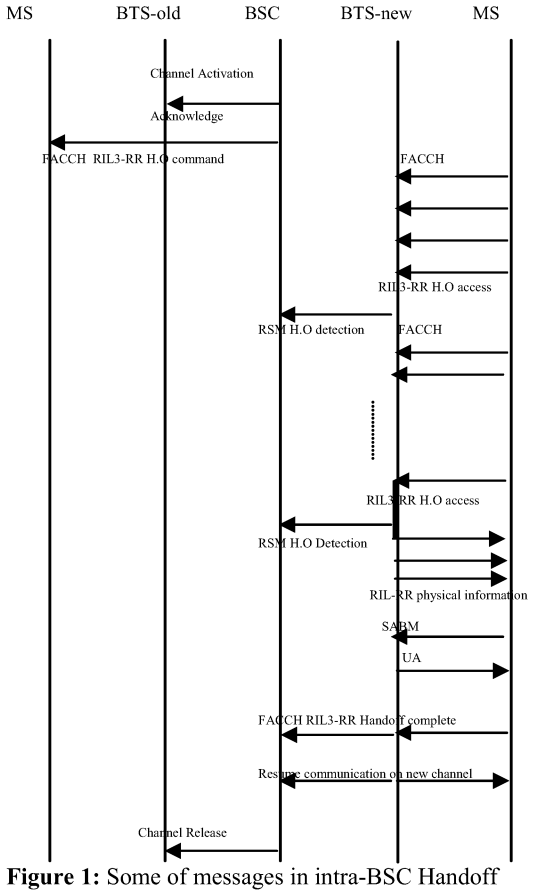


Figure 1: Some of messages in intra-BSC Handoff

Suppose that the time of holding two channels for one handoff is known as wasted time due to handoff ($T_{H.O}$). The calculation of wasted time due to handoff ($T_{H.O}$) in Intra-BSC handoff in second generation (2G) mobile communication is calculated as shown in Table 1.

Table 1: Calculation of the time of holding two channels

Message	Interface	Minimum Length of Layer 3 message "Note 1"	Length of L3 message "Note 2"	Length of L2 message "Note 3"	Total transmitted bytes	Min. time
Channel Activation Acknowle-dge	Abis	7 bytes	7 bytes	7 + 8 = 15 bytes	15 bytes	14 bytes *125µs + 3.9µs = 1.7539ms
Handoff command	Abis-Um	10 bytes	10 byes + 9 bytes = 19 bytes	19 + 8 = 27 bytes	27 bytes	26 bytes *125µs + 3.9µs = 3.2539ms
synchro-nous Handoff	1) Abis					
Asynchron-ous Handoff	1) Abis	9 bytes	9 byes + 9 bytes = 18 bytes	18 + 8 = 26 bytes	26 bytes	25 bytes *125µs + 3.9µs = 3.1289ms
synchro-nous Handoff	2) Um	10 bytes	10 bytes	10 byes + 11 bytes + 3 bytes = 24 bytes	456 bits = 57 bytes	7 bursts* 4.615ms + 0.577 = 32.882ms
Asynchron-ous Handoff	2) Um	9 bytes	9 bytes	9 byes + 11 bytes + 3 bytes = 23 bytes		
Handoff access burst	Um	1 bytes	1 bytes	1 bytes	36 bits	1 bursts* 0.577ms = 0.577 ms
synchro-nous Handoff						• 1.154 + 3 * 4.615 = + 0.577 = 15.576 ms
Asynchron-ous Handoff						1.154 + 4 * 4.615 = + 0.577 = 20.191 ms
Physical informa-tion	Um	3 bytes	3 bytes	3 byes + 17 bytes + 3 bytes = 23 bytes	456 bits = 57 bytes	•• 1.154 + (7 bursts* 4.615ms) + 0.577ms = 34.036 ms
••• SABM	Um	3 bytes	3 bytes	3 byes + 17 bytes + 3 bytes = 23 bytes	456 bits = 57 bytes	1.154 + (7 bursts* 4.615ms + 0.577 = 34.036 ms
•••• UA	Um	3 bytes	3 bytes	3 byes + 17 bytes + 3 bytes = 23 bytes	456 bits = 57 bytes	1.154 + (7 bursts* 4.615ms)+ 0.577 = 34.036 ms
Handoff complete	Um - Abis 1) Um	3 bytes	3 bytes	3 byes + 17 bytes + 3 bytes = 23 bytes	456 bits = 57 bytes	1.154 + (7 bursts* 4.615ms)+ 0.577 = 34.036 ms
	Um - Abis 2) Abis	3 bytes	3 byes + 9 bytes = 12 bytes	12 byes + 8 bytes = 20 bytes	20 bytes	19 bytes * 125µs + 3.9 µs = 2.3789 ms
Channel Release	Abis	4 bytes	4 bytes	4 bytes + 8 bytes =12 bytes	12 bytes	11 bytes * 125µs + 3.9µ s = 1.3789 ms
Time of holding two channel in Asynchronous Handoff						**197.8576 ms**
Time of holding two channel in synchronous Handoff						**91. 2596ms**

Note 1: -Layer 3 message: - the signalling messages that will be created in layer 3 known as Layer-message.

Note 2: -L3 message: - the messages that is forwarded to layer 2.

Note 3: -L2 message: - the message that is forwarded to layer 1.

● The BTS transmits three time slots later than the mobile station =1.154 ms.

●● When sending the physical information message, the network starts timer T3105. If this timer times out before the reception of a correctly decoded layer 2 frame (SABM), the network repeats the physical information message and restarts timer T3105. To calculate minimum wasted time, suppose that this message will be transmitted one time.

●●● The information field is permitted with the SABM command; layer 3 will indicate when an information field is to be included (Handoff complete in the Handoff procedure). (TS GSM 04.06)

●●●● UA frame carries the same information field contents as the SABM frame from the MS.

From Table (1), it can be seen that the time of holding two channels is equal to 197.8576 ms for asynchronous handoff and 91.1346 ms for synchronous handoff.

Handoffs in GPRS are strongly aligned to GSM handoffs, so the wasted time due to handoff ($T_{H.O}$) in GPRS can be calculated in a similar method.

MATHEMATICAL MODEL

There are many parameters affecting the probability of handoffs depends on it such as cell size, user velocity, direction of motion and call holding time.

In the probability theory, it is known that the probability of n occurrence of any event X is less than the probability of m occurrence of X if m < n. Let the probability that a call makes a handoff after it has make 0, 1, 2, ……., n previously successful handoffs is Pn , then Pn decay as exponential as shown in Figure (2) (Papoulis 1991).

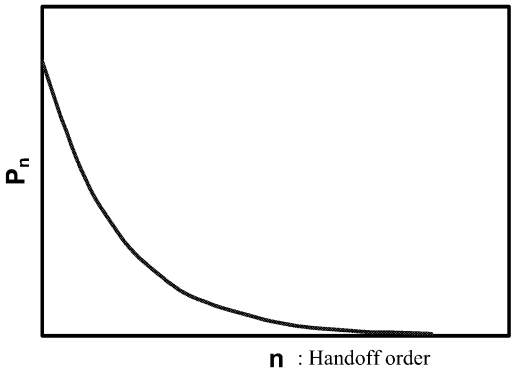


Figure 2: Probability of handoffs versus handoff order

The total delay time per user can be calculated from state machine model as shown below Figure (3).

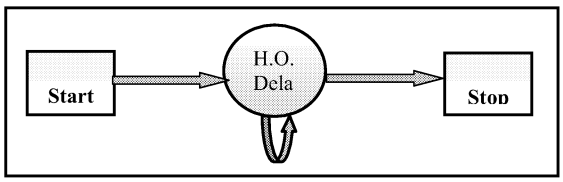


Figure 3: State machine for calculation the total delay

This is a probabilistic model. The transfer function for this model using the standard notation of graph theory is given by:

$$\frac{O}{I} = P_1 D + P_2 D^2 + P_3 D^3 + \ldots\ldots\ldots \quad (1)$$

Where:

Pn: the probability of handoffs for the nth times.

D: the average number of handoff in a user call (1/sec) and

D^n: represents n repetitions of D.

So the total delay is equal to: -

$$D_{total} = \sum_{i=1}^{\infty} P_i D^i \quad (2)$$

The total wasted time due to handoffs per user is

$$T_{wasted} = D_{total} \times T_{H.O} \times \lambda \quad (3)$$

Where: λ is the call holding time (sec)

The cell capacity reduction will be as follows:

$$C_{\downarrow} = \frac{T_{wasted}}{\lambda} \times 100\ \% \quad (4)$$

This model is based on calculating the probability of Handoffs. It depends on many factors in the system such as cell size, human behavior and sectorization. It is found that it is hard to obtain closed form formulas for Handoffs effect over capacity. Therefore, from simulation, the number of handoffs versus different factors (different cell sizes, average velocities of users and average call holding times) was found.

SIMULATION

The first simulation model used considers 49 cells system without sectorization and with 2 percent of average blocking rate.

The number of channels per cell is assumed to be 54 traffic channels (Time-slots) and the arrival process of new calls in a cell is assumed to be Poisson process with rate (u). At call initiation, a mobile is assigned a random initial position from a uniform distribution over the cell area. The call is also assigned a random total duration (call holding time) from Poisson distribution with rate (λ). The velocity of user is assumed to have an exponential distribution with a mean of (v) (Kulavaratharasah and Aghvami 1999, Lin and Chlamtac 1999). It is also assumed that the direction of motion follows uniform distribution.

In the second simulation program, we consider 49 cells with sectorization (3 cell site) and similar assumptions are made as in the non-sectorized simulation.

The algorithm simulation program is described as follows:

- Set several values for cell radius, average velocity of the user and call holding time.
- For each cell, assign users, their velocity and call holding time by using the values that are selected from step one and also assign their location and their direction of motion.
- Start simulation for some period of time, and in each one second update the location of each user, the call holding remaining time, the Home cell number, then search for users that makes a handoffs. Also, in each loop, remove users who finished their call and add new users.
- After the simulation is ended, we determine the number of handoffs and plot it versus velocity, call holding time and cell radius.

RESULTS AND CONCLUSIONS

Figures 4-7 shows the simulation results of the cellular system with non-sectorized cells considering the following factors: cell size, velocity and call holding time.

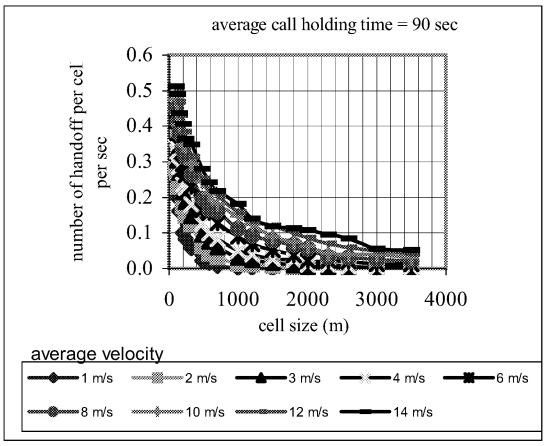


Figure 4: Handoffs versus cell size in Non-Sectorized Cells

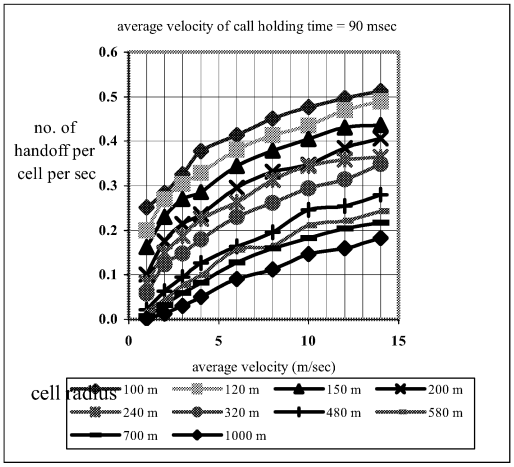


Figure 5: Handoffs versus average velocity for Microcells cell size in Non-Sectorized Cells.

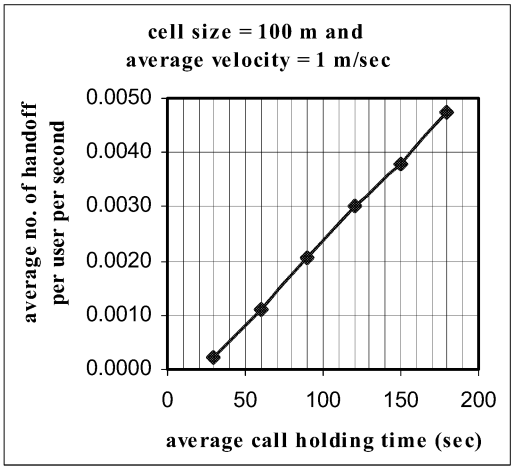


Figure 6: Handoffs versus average call holding time in Non-Sectorized Cells for R=100 m and v=1m/sec

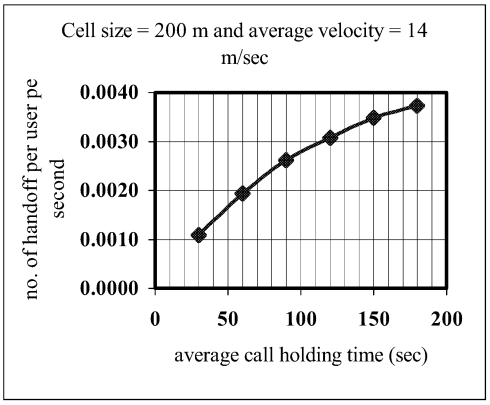


Figure 7: Handoffs versus average call holding time in Non-Sectorized Cells for R=200 m and v=14m/sec

From these results, it can be conclude that the probability of handoffs increases as the cell size decreases as one could easily predict and handoffs probability increases as average velocity of the user increases this is because the probability of a mobile station crossing the border of one cell into another will increase. The probability of handoffs increases as average call holding time increases specially for small cell size and as user mobility increases. It is obvious that when the call takes too long, there is a high possibility that a user might move and change location and this effect becomes more significant if his motion is high and the cell that he lives in is small.

Figures 8-11 shows the simulation results of the cellular system with sectorized cells assuming three radios per sector considering the following factors: cell size, velocity and call holding time.

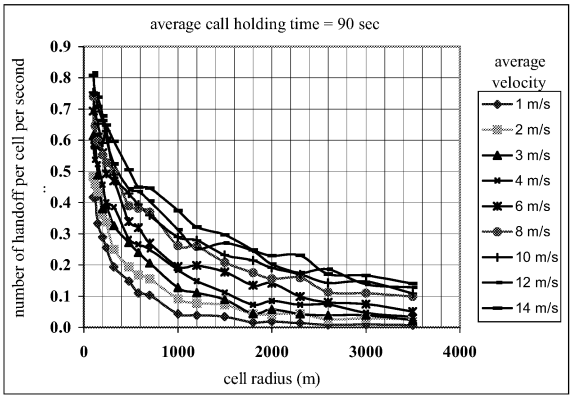


Figure 8: Handoffs versus cell size in Sectorized Cells.

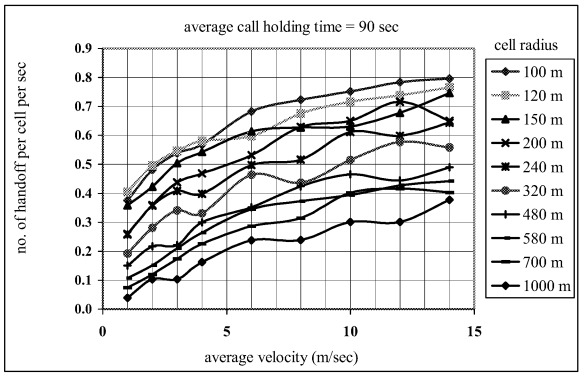


Figure 9: Handoffs versus average velocity for Microcells cell size in Sectorized Cells.

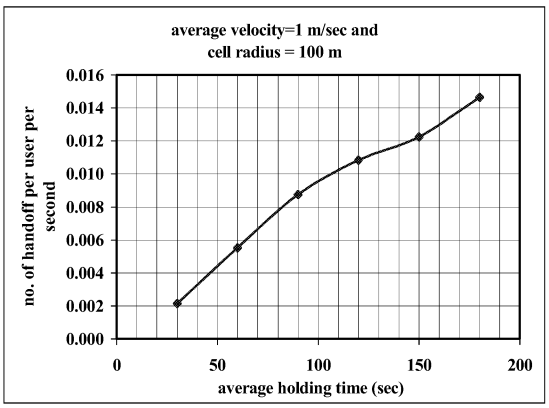


Figure 10: Handoffs versus average call holding time in Sectorized Cells for R=100 m and v=1m/sec

We can observe from these simulation results that the handoffs probability decreases as cell size increases; as in non-sectorization cells but the probability of handoffs for sectorized cells is higher than that for non-sectorized cells for specific average velocity and average call holding time. The explanation of this observation is that the sectorization technique split a cell into a number of sub-cells, each sub-cell behaves as an independent cell, and so smaller cells will result in sectorization technique. And the handoffs probability increases as average velocity of the user increases; as in non-sectorization cells but the probability of handoffs for sectorized cells is higher than for non-sectorized cells, due to the fact that users in sectorized cells may make handoffs even if their positions are near the center of the cell.

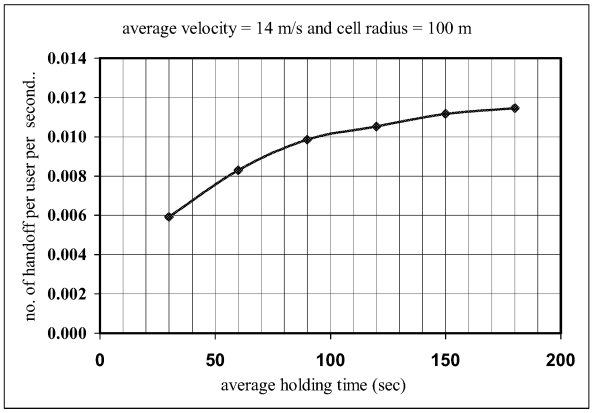


Figure 11: Handoffs versus average call holding time in Sectorized Cells for R=1000 m and v=14m/sec

From the simulation results discussed in this paper, the capacity reduction is calculated at the 100m-cell radius and v =14 m/sec for the sectorized case in case asynchronous handoff, which have higher wasted time than synchronous handoff, and it is equal to 16%.

A mathematical model was proposed from numerical curve fitting for the data we obtained and it is given by:

$$C_{\downarrow} = \alpha\,[a + b\ln(\lambda)]\,e^{(f\ln(v)+d\sqrt{R})} \quad (5)$$

Where:

a =- 0.0055, b = 0.0042, f = 0.29, d = -0.06, α = 8.9159

λ = average call holding time in (sec)
v = average user velocity in (m/sec)
R = cell radius in (m)

Comparing with the mathematical model, we obtain:

$$T_{H.O.} \sum_{i=1}^{\infty} P_i D^i = \alpha [a + b \ln(\lambda)] e^{(f \ln(v) + d\sqrt{R})} \quad (6)$$

Therefore, from simulation and mathematical model, the probability of handoff for the ith times (p_i) equation was proposed as a function of v, R and λ.

References

Chung-Ming Huang, Chao-Hsien Lee, Chung-Yi Lai, and Yuan-Tse Yu, 2004, A Location-aware Layer 7 Proxy Handoff Mechanism over the Mobile Network Environment, Proceedings of the 18th International Conference on Advanced Information Networking and Application (AINA'04) 0-7695-2051-0/04 © 2004 IEEE.

Ericsson and Telia, 1998, Understanding Telecommunication 2, Ericsson Telecom, Telia and Stusdentlitterature, Sweden.

Guerrero, L.O. and Aghvami, A.H., 1999, A Prioritized Handoff Dynamic Channel Allocation Strategy of PCS, IEEE Trans. Veh. Technol, 48 (4): 1203-1215.

J. Wang, Q-A. Zeng, and D. P. Agrawal, "Performance Analysis for Preemptive and Priority Reservation Handoff Scheme in Integrated Service Based Wireless Mobile Networks" IEEE Transactions on Mobile Computing, Vol. 2, No. 1, January-March 2003, pp. 65-75.

Kulavaratharasah, M.D. and Aghvami, A.H, 1999,Teletraffic Performance Evaluation of Microcellular personal communication networks (PNC's) with prioritized handoff procedures, IEEE Trans.Veh. Technol. Commun 48 (1): 137-152.

Lee, W.C.Y., 1996, Mobile Cellular Telecommunication Systems, McGraw Hill, New York.

Lin, Y.B., and Chlamtac, I., 1999, Effect of Erlang Call Holding Times on PCS call Completion, IEEE Trans. Vech. Technol, 48 (3): 815-823.

Qingan Zeng, and Dharma P. Agrawal, "Modeling and Efficient Handling of Handoffs in Integrated Wireless Mobile Networks," IEEE Transactions on Vehicular Technology, Vol. 51, No. 6, November 2002, pp. 1469-1478.

Papoulis, A., 1991, Probability, Random Variables, and Stochastic Processes, third edition. McGraw-Hill. Inc. New York.

Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI) 2000, GSM 04.06 (ETS 300 938): "Digital cellular telecommunications system; Mobile Station - Base Station System (MS - BSS) interface Data Link (DL) layer specification", Fifth Edition.

Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI) 1995, GSM 04.07 (ETS 300 556): "Digital cellular telecommunications system; Mobile radio interface signaling layer 3; General aspects".

Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI) 1997, GSM 04.08 (ETS 300 940): "Digital cellular telecommunications system; Mobile radio interface layer 3 specification".

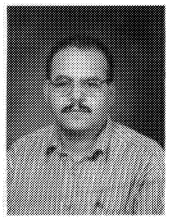
Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI) 1998, GSM 08.58: "Digital cellular telecommunications system; Base Station Controller-Base Transceiver Station (BSC-BTS) interface; Layer 3 specification".

Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI) 1995, GSM 03.09: "Digital cellular telecommunications system; Handover procedures".

AUTHOR BIOGRAPHY



Nada Khatib received her BSc degree in electronics and communication engineering from the Baghdad university, Iraq in 1996 and MSc degree in communication engineering from the Jordan university, Jordan in 2001. She is currently a lecturer in the Department of communications and Electronics Engineering at Philadelphia University, Jordan. Her research interests include Digital communication and cellular communication systems.



Jamal Rahhal received his BSc degree from the University of Jordan, Jordan 1989, the MSc and Ph.D degree from Illinois Institute of Technology, Chicago IL, in 1993, 1996 respectively. He is with the Faculty of Electrical Engineering in the University of Jordan since 1996. His research interests including CDMA cellular communication systems, wireless communication systems, wavelet packet transform for multiple access techniques, array processing, pattern recognition, and applications of Neural Networks and Fuzzy Systems.

Computational Aspects to Calculate the Interference between the High Voltage Transmitters and Communication Cables

GHADA M. AMER
Electrical & power engineering department
Higher Institute of Technology
Benha, Egypt
dr_ghada11@hotmail.com, dr_ghada11@yahoo.com

Abstract: - Transmission lines are considered one of the major sources of magnetic field. The EMF interference with communication lines located in utility corridors is a real and serious problem, which can place both operator safety and communication line integrity at risk. This paper discussed the effect of the inductive and ohmic interference voltage of overhead transmission lines in short circuit conditions on communication circuits runs parallel with power networks

Key Words: - EMF interference, the inductive interference, ohmic interference, H.V.T.L, communication cables

1. INTRODUCTION

Installing communication lines in energy utility corridors containing high voltage AC transmission lines subjects the communication lines to induced AC voltages. This can be caused by an imbalance in the transmission system and by high voltages near transmission tower grounding systems resulting from lightning strikes and phase to ground faults. When a long-term induced AC voltage exists on a communication line, it can be dangerous and potentially life threatening for operations personal to touch communication line or appurtenances [1]. In addition the metallic cheese of communication cables corrosion also can results from AC discharge. Recommendations on permissible voltages calculations and protection methods have been issued by CCITT [2] in its recommendation for protection of telecommunication lines against adverse influences of power lines, by the IEC and by DIN VDE [2]. According to these standard specifications communication lines and cables induced voltages must not exceed the following values between conductor and earth: -

i. 60 - 65 V in long time interface communication circuits, which are not, terminated with isolating transformers and voltage arrestors.
ii. 150 – 250 V long-time interference in communication circuits, which are terminated with isolating transformers or other special conditions, are provided.
iii. 300 – 500 V in short time interface ≤ 0.5 second.

And it is apparent that when the ground fault current is very high even a grounding rod having low grounding resistance may produce potential gradient on the ground which may be dangerous to human and animal life [3], and it may increase the induced EMF in the neighboring communication lines. For providing proper safety it is essential to check and keep in limits the step and touch potentials by decreasing the resistance still further or by using more than one grounding rod [4]. When the earth resistivity is high and the driven rods are not adequate so the counter poise conductor can be used [5-7]. In this paper calculations are carried out to investigate the inductive and ohmic interference communication voltages by using EMTP. The calculations are done when the high voltage transmission line is exposed to different earthed short circuit cases; the effects of soil Resistivity as well as the horizontal distance between the two circuits are taken into consideration. The

effect of using the counter poise between high voltage transmission line towers is also investigated.

2. MODELING of High Voltage Transmission Lines by Using EMTP: -

Electro Magnetic Transient Program EMTP is highly accurate digital simulation program for high voltage three-phase transmission line simulation. EMTP is used to produce voltage and current waveforms for different type of faults. The transmission line components, which are modelled by the programs, included: (a) resistance, inductance and capacitance. (b) Travelling wave models to represent overhead lines or cables. (c) Ideal current and voltage simulation. The inputs to the program are: The length of line to be simulated, the model parameters data, the length of time to be simulated, calculation of time step (sampling time) as well as the desired output. The transmission line is represented by nine π models scaled to the parameters of 500 kV. The transmission line parameters are calculated from line geometrical characteristics. The calculated parameters are expressed as series impedance and shunt capacitance per unit length. As follows: R_{11}= 2.243495e-1, X_{11}=9.631423e-1, R_{21}= 1.808330e-1 X_{21}=4.462796e-1, R_{22}= 1.918223e-1, X_{22}=9.840801e-1, R_{31}= 1.808330e-1, X_{31}=4.462796e-1, R_{32}= 1.658927e-1 X_{32}=4.426757e-1, R_{33}= 1.918223e-1, X_{33}=9.840801e-1 Where R_{11}, R_{22}, R_{33}, X_{11}, X_{22}, and X_{33} are the self-resistance and reactance for the phase conductors. And R_{12}, R_{21}, R_{13}, R_{31}, R_{23}, R_{32}, and X_{12}, X_{21}, X_{13}, X_{31}, X_{23}, X_{32} are the mutual resistances and reactances between the phases. Transmission line used single parallel continuous counter poise Fig 1.



Fig 1 single parallel continuous counter poise.

Overhead earth wire interconnecting the earth electrodes at various towers to provides a number of parallel paths to the ground fault current and this also provides a shield against lightning strokes.

3. Intereference Analysis

Interference between transmission systems and communication can be of three kinds [8]: *capacitive interference*, which, occurs in the immediate vicinity of the overhead power line when the communication circuit is well insulated from the ground. The communication circuit picks up a voltage relative to earth, which is proportional to the voltage of transmission line. *Resistive or ohmic interference* occur when lightning strikes a transmission structure, or when there is a phase-ground fault. When this occurs, a large voltage cone is created around the ground equipment. If communication cable is located with this area, voltage can get onto the communication circuit in the area within the voltage cone through coating. *Electromagnetic or inductive interference* occurs when there is extended and close parallel routing with three phase high voltage AC transmission lines. The voltage is due to any phase imbalance in the lines.

3.1 Inductive interference

In case of phase to ground faults, and in Transmission line used single parallel continuous counter poise. The ohmic and capacitive interference will be neglectable with respect to the inductive interference. By the magnetic field formed by the circuit phase conductor earth an EMF is induced in the communication cable conductor to earth. The proportionality factor is the mutual inductance M_E. The value of mutual inductance depends upon the dimensions of the two circuits and their position relative to one another. If these cross at right angles, then the mutual inductance M_E=0, then M_E is directly proportional to the length of the parallel run. In the event of an earth short circuit on a power line, an EMF E_i is induced in the conductors of the neighboring communications cable, which is

$$E_i = \omega . M_E . L . I . r \quad (1)$$

Where E_i induced EMF in volts, ω frequency s^{-1}, M_E mutual inductance per

unit length between the influenced and influencing cable (or line) in H/m, L is the length of parallel run in meter, I is the interface inducing current (e.g. earth fault) in A [2], and r is overall reduction factor (product of several reduction factors). Such reduction factors are: reduction factor for metallic covering of power cable, reduction factor for the metallic covering communications cable, reduction factors for compensating conductor as in Fig. 2.

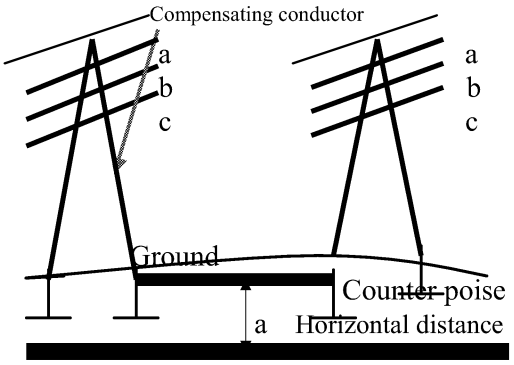


Fig 2 Model of H.V.T.L. with compensating conductor and counter poise conductor.

The mutual inductance M_E is dependant upon the distance between the influenced and influencing conductors, the frequency and the specific earth resistance. The mutual inductance M_E can be calculated from the following formula [9]:

$$M_E = X.Y^a.a^z \qquad (2)$$

Where X, Y, and Z are variable coefficients depend on the specific earth resistance, 'a' is communication cable and power line converges, the calculation must be made using a mean distance 'a' is calculated from the values a_1 and a_2 at beginning and end of the convergence.

$$a = \sqrt{a_1.a_2} \qquad (3)$$

It is found that [9] ρ = 5000 Ωm rocky ground X= 1.9093137, Y= 0.99782865 and Z=-0.090552262, for ρ = 50 Ωm loam or cultivated ground X= 1.2882006, Y= 0.99532428 and Z=-0.133556 and for ρ = 0.1 Ωm urban areas X= 0.68741049, Y= 0.95819727 and Z=-0.24383596.

3.2 Ohmic Interference

In case of phase to ground faults, and in transmission line without counter poise conductor. The inductive and capacitive interference will be neglectable with respect to the ohmic interference [10]. In the event of a S.C. to earth the voltage drop which develops on the earthing equipment, due to earth impedance Z_E, appears between the conductor of the communication line and earth as shown in Fig. 3.

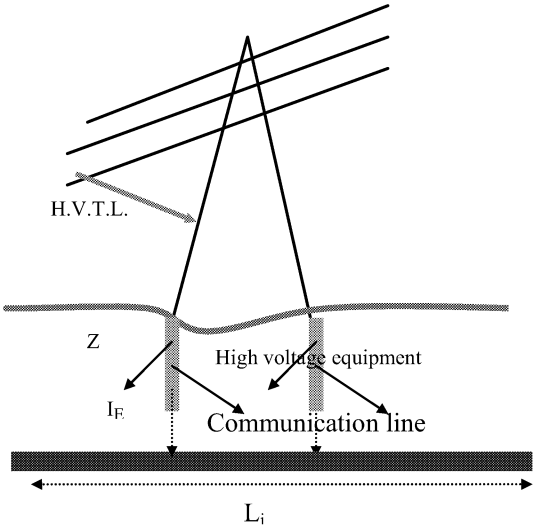


Fig. 3. Ohmic interference on communication line

The ohmic interference voltage is therefore:

$$E_{\Omega k} = I_E.Z_E.r_{ku} \qquad (4)$$

Where $E_{\Omega k}$ is the interference voltage due to ohmic coupling in V I_E is a part of earth current which flows from earthing network to earth in A. This current is never greater than the product of the earth fault current and the largest reduction factor of one of the power conductors of the faulty network [2]. Z_E is the impedance to earth of earthing equipment in Ω and r_{ku} is the reduction factor of the communication cables. The ohmic longitudinal voltage per unit length for determination of the effective reduction factor can be calculated, in this case, with the following formula:

$$E_i' = \frac{I_E.Z_E}{I_{is}} \qquad (5)$$

Where E_i' is ohmic longitudinal voltage per unit length in V/m, I_E is the part of earth current which flows through the earthing arrangement to ground in A and I_{is} is the insulated cable length in m.

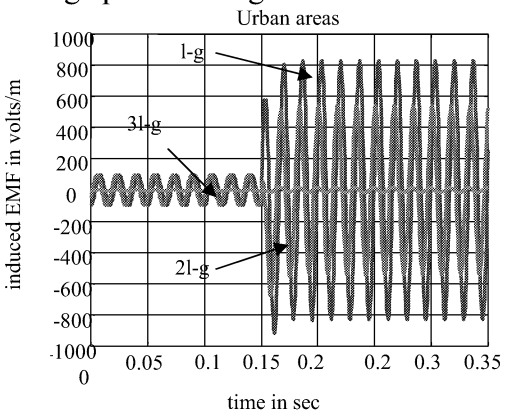
4. Results and Discussion

4.1 Effect of inductive interference

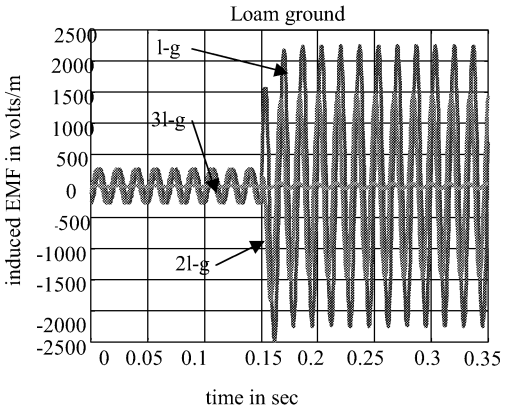
The induced EMF of high voltage transmission lines with counter poise on the communication cables is calculated in different ground fault types. The horizontal distance as given in Fig 2 between the transmission line counter poise and communication line is taken 5 m. The length of parallel run is 1000m, f=50 Hz, the overall reduction factor is choose to be 1, and the interface inducing current (e.g. earth fault) is the output of EMTP. The mutual inductance per unit length is calculated from Eq. (2) by changing the specific earth resistance. By using Matlab program (M-file) Eq. (1) is used to get the inductive interference in volts as a function of time. Fig 4 shows the inductive interference voltages for the three fault types (a) in case of ρ=0.1 Ωm (urban area), (b) in case of ρ=50 Ωm (loam ground) and (c) in case of ρ=5000 Ωm (rocky ground). From Fig 4, it can be observed that the inductive interference in communication line is increases with increasing the specific earth resistance as it is expected. From this figure it is noticed also that the 3-phase to ground fault gives the least inductive interference in the communication line and this is in agreement with [10,11], and it is lesser than 18.9797 V/m in case of ρ=0.1 Ωm (urban area), and phase to ground fault gives the most serious case where inductive interference in case of ρ=0.1 Ωm (urban area) is 831.9 V/m.

4.2 Effect of ohmic interference

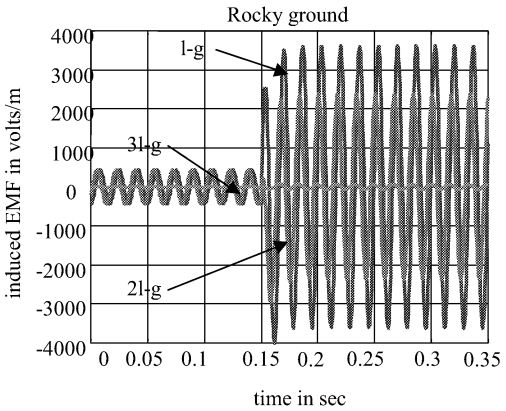
The interference is from transmission line without counter poise conductor with grounding system as shown in Fig. 3 the interference is calculated from equation 4. The insulated cable length is 1000 m and I_E is 70% of I_f [2]. Fig. 5 (a) shows the ohmic longitudinal voltage per unit length in V/m.



(a) ρ=0.1 Ωm (urban area)



(b) ρ=50 Ωm (loam ground)



(c) ρ=5000 Ωm (rocky ground)

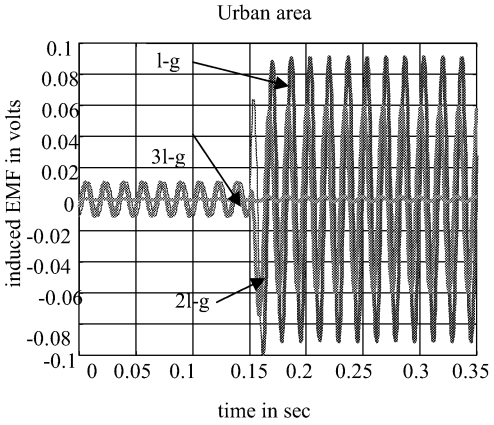
Fig. 4 the inductive interference in volts/m for all ground fault types

The equivalent resistance of the tower earthing system Z_E is calculated by Eq. (6).

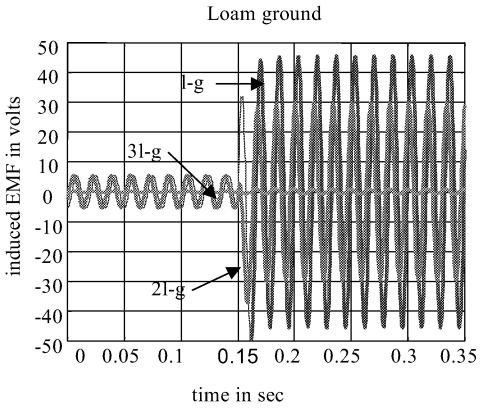
$$Z_E = \frac{\rho}{2\pi L n} \ln\left(\frac{8L}{d} - 1\right) \qquad (6)$$

Where n is the number of electrode, ρ in the specific earth resistance, L is the

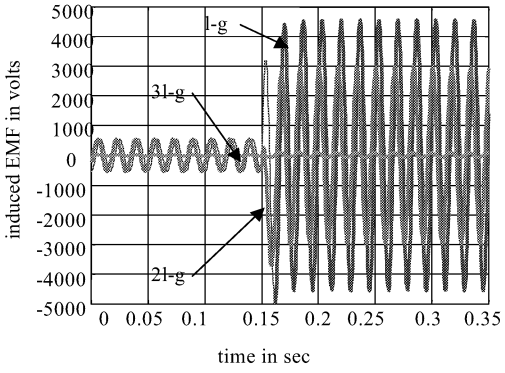
length of the rod and d is the diameter. L is taken constant and equal 1m and d is taking 2.5 cm. The impedance to earth of earthing equipment Z_E is changing by changes the specific earth resistance. By using Matlab program (M-file) Eq. (5) is used to obtain the ohmic interference in volt/m as a function of time. Fig 5 shows the ohmic interference for three fault types (a) in case of ρ=0.1 Ωm (urban area), (b) in case of ρ=50 Ωm (loam ground) and (c) in case of ρ=5000 Ωm (rocky ground).



(a) ρ=0.1 Ωm (urban area)



(b) ρ=50 Ωm (loam ground)



(c) ρ=5000 Ωm (rocky ground)

Fig. 5 ohmic interference in volt/m for all ground fault types for all fault types.

From Fig 5 it is observed that the least value for the ohmic interference is at 3 phase to ground fault at the urban area and its equal 0.0021 V/m, and the highest value for this type of the interference is for phase to ground fault at rocky ground when ρ=5000 Ωm, the ohmic interference in this case is equal 4561.4 V/m. Fig. 6 shows the ohmic and inductive interference with respect to the specific earth resistance.

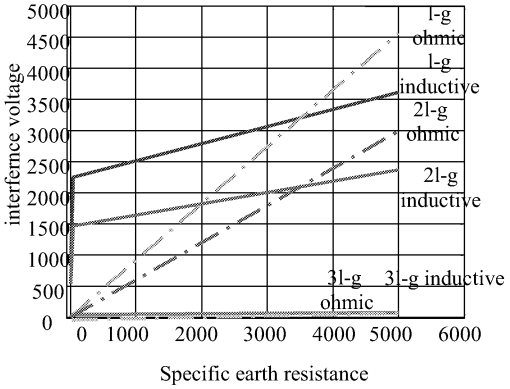


Fig. 6 Ohmic and inductive interference with respect to the specific earth resistance.

It's observed from Fig. 6 that the ohmic interference voltage is less than the inductive interference at ρ=0.1 Ωm and voltage at ρ=50 Ωm for all ground fault types. But at ρ=5000 Ωm the inductive interference voltage is lesser than ohmic interference. In general it's noticed from Fig.6 that at low specific earth resistance the inductive interference voltages between high voltage and communication cables are higher than the ohmic interference voltages. At higher specific earth resistance the rate of increase of ohmic interference is higher than the inductive interference and this leads to high values of ohmic interference between the communication circuit and power circuit.

5. Conclusion

- Novel technique is used to calculate ohmic and inductive interference voltages between the high voltage transmission lines and communication cables as a function of time in different ground faulty cases.

- The inductive and ohmic interference in communication line increases with increasing the specific earth resistance.
- At low specific earth resistance the inductive interference voltages between high voltage and communication cables are higher than the ohmic interference voltages.
- At higher specific earth resistance the rate of increase of ohmic interference is higher than the inductive interference and this leads to high values of ohmic interference between the communication circuit and power circuit.

References:

[1] Anderson L. 2000. *EMF responses in farm animals*. In: Proceedings of International Seminar on Effects of EMFs on the Living Environment, October 1999, Ismaning Germany. Oberschleissheim: ICNIRP, 51-56.

[2] Power Cables and their Application, part 1, Lothar Heinhold, 3rd revised edition, 1990.

[3] I. LINGVAY, Carmen LINGVAY, F. STOIAN, Corina BABUTANU, "Solid State Device for Electric Safety of Metallic Structures and Electronic Equipment" – the 17th International Symposium on Naval and Marine Education, May, 24-26 2001, Section 2, pp. 278-282.

[4] J. Peters, "Community Exposures to Electric and Magnetic Fields and Possible Health Effects", presented at the EPRI sponsored seminar "Epidemiological Studies of Electromagnetic Field Exposure", Minneapolis, Minn., 1987.

[5] Horizontal Strip Electrodes for Lowering Impedance to Ground", Earthlink Application.

[6]http://www.graphitesales.com/advan.htm.

[7] Electric Power Distribution" by T. A. Short, edited by CRC Press (2004).

[8]"Guide Concerning Interference of High Voltage AC Power System on Metallic Pipelines" CIGRE Study Committee, July 1992.

[9] Ghada M. Amer "Effect of Soil Resistivity on Electromagnetic Induced Voltage of Communication Cables Installed Parallel to High Voltage Transmitters", WEAS conference, Kourfu 2005.

[10] Y. Li, F. P. Dawalibi, and J. Ma, "Electromagnetic Interfernce Caused by a Power System Network and a Neighboring Pipeline" Proceeding of the 62nd Annual Meeting of the American Power Conference, Chicago, April 10-12, 2000.

[11] G. M. Amer 'Novel technique to calculate the effect of electromagnetic field of H.V.T.L. on metallic pipeline by using EMTP', 18th CIRED conference, Turin 2005.

IMAGE AND DATA MANIPULATION SIMULATION

Image segmentation based on games theory approach Application to melanoma pattern extraction

Amin Kallel and Imene Khanfir
National Engineering School of Sfax (ENIS)
B.P.W, 3038 Sfax, Tunisia
Research Unit on Intelligent Control, design & Optimisation of complex Systems (ICOS)
amin.kallel@enis.rnu.tn

Abstract
The area growth algorithms consist in traversing the image to look for the core of a new area, called germ, among its pixels that are not yet classified. The growth of the area corresponding to this germ is done by the addition of some of its close pixels which present colorimetric resemblances. Thus, the results of such approaches depend closely on the way according to which the germs are selected. A method of dynamic segmentation is proposed to overcome this problem. It is deduced from the general concepts of the game theory and consists in representing the core of a new area by a whole set of pixels and not by only one germ.

1 Introduction

The classification of the nævic lesions is one of the most significant current stakes of the dermatology and the medical imagery. Indeed we can note a growing interest for the development of data-processing tools to nævic lesion diagnosis assistance. Such a tool would consist initially, to extract the lesion from its background to allow its analysis thereafter. This corresponds at the stage of segmentation [3].

The segmentation, aims to share the image in homogeneous areas. Two approaches of segmentation are distinguished: The segmentation by contour detection [8] and the segmentation by area extraction [4]. The latter, using homogeneity criteria can separate two areas having different characteristics even if the border between them is not clear. This case being the nævic-lesion image one, the segmentation by the area extraction approach is more adapted to our application.

The development of a segmentation algorithm of medical images of the nævic-lesions consists in extracting as precisely as possible the lesions from their background. Such algorithms were built in order to gather the pixels corresponding to the lesion in only one area and the remainder of the pixels in another area normally representing the healthy skin [8].

By examining the game theory, we note that there is an analogy between its step related to the constitution of the coalitions and the methods of automatic segmentation having as objective the search for a distribution of an image in homogeneous areas. Then, we wanted to benefit from this analogy to conceive a method of segmentation well adapted to the extraction of a nævic-lesion from its background.

In this context, this paper is organized as follows: in the second section, we briefly describe some segmentation methods that were deployed by other researchers for the extraction of an object from its background. We present in the third section general information on the game theory while emphasizing the classification of pixels in distinct areas. We propose the segmentation method inspired from this theory in the fourth section. In the fifth section, the results of the dermatological lesion extraction, ensured by this segmentation method, are presented and commented. We also present the results of extraction assured by two other methods in order to compare them with our results and to situate our method compared to existing ones. Lastly, we end up by concluding and giving our perspectives.

2 Segmentation by region extraction

The areas image segmentation is assured either by a "pixel classification" approach, or by an "area growth" approach [4]. The principal difference between these two approaches consists in the way of the use of the grey level space and relations existing between the pixels.

✓ The methods of pixel classification initially determine a partition of the grey level space. The connexity relations are then useful to determine the areas. These methods generally use techniques of grey level histogram thresholding [9]. Several techniques of histogram thresholding [10-16] exist in the literature. These methods were very often developed to treat the particular case of the segmentation in two classes: it is the

conversion to a binary image. The two classes generally represent the object and its background (in our case it is the lesion and the healthy skin). The pixels with values exceeding the threshold belong to the image background; the others belong to the object (cf figure 1).

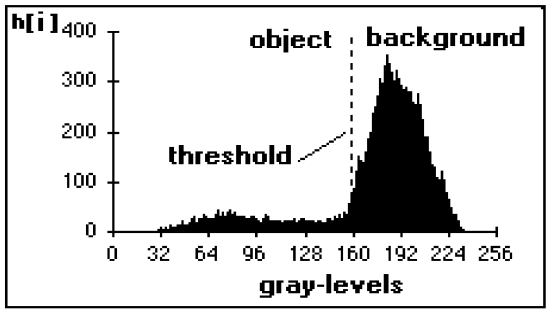


Figure 1 : Bimodal histogram.

Five algorithms were tested for the segmentation of the dermatological lesion images [5]: the Moments algorithm [10], the Entropy algorithm [11], the iterative Intermeans algorithm [12-13], the Intermeans algorithm [14] and the Triangle algorithm [15].

✓ The methods of area growth simultaneously use the two types of information: the gray level information and the relations existing between the pixels. They consist in enlarging an area according to an homogeneity and vicinity criteria, starting from one selected pixel [16]. Accordingly, several algorithms were developed:

- Trémeau and Borel [17] propose an algorithm of segmentation by pixel aggregation, which proceeds by a sequential sweeping of the image and regards the first pixel as a germ. They then tried to enlarge this germ as much as possible by incorporating the nearby pixels.
- Moghaddamzadeh and Bourbakis [18] proposed a cooperative method of area growth and contour detection. In a first stage, the contour pixels are detected by an intensity gradient thresholding. In a second stage, they only considered pixels which belong to uniform areas to build the regions by pixel aggregation.

The presented approaches provide results depending on the image sweeping direction. To avoid this problem, Shafarenko and Petrou [19] proposed a method of topographic area enlarging which is founded on the watershed.

3 Games theory

While referring to the general concepts of the game theory [6], we deduce that the systems of alliances in the co-operative games constitute pretopologies on the whole of the players. We also deduce that at the end of a game, such systems constitute a partition on the whole of the players, similar to an optimal partition in the field of automatic classification [2].

A "game" is a situation in which several individuals (that can be pixels), have to make some decisions on which depend the results that are related to the whole of them. The games of bridge or poker constitute examples of such situations. But the definition is sufficiently general to cover much diversified phenomena such as military conflicts, political party conflicts, economic competes... and also automatic classification problems or image segmentation.

We distinguish two situations which can characterize a game: the co-operation or the fight. Indeed, individuals confronted in a game can have on certain points, opposite interests.

In general, a game does not proceed instantaneously. Infect, on the one hand the decision effectively made by each player can be the object of a bargaining during which the positions can change. On the other hand, the game can be a succession of elementary decisions made by each player and having for object a single consequence. Such a game is called a dynamic game.

The games, characterized by situations of pure fight, are only particular cases of the games. So, we rarely find their trace in the practical life. For this reason, the method that we propose for the segmentation of the nævic lesion images, obeys to the principles of the co-operative games.

We call co-operative games (partially or completely) the games in which the players can be combined by various ways to obtain more than they could obtain while remainding alone.

We admit here that the stake can be characterized by a measurable dimension (equivalent to coins) that the players will try to get while gathering. Each group thus formed is called coalition (area for our application). It will be noted that an isolated player is always a particular case of coalition.

4 Proposed method

The usual methods of area enlarging segmentation, rest only on one predicate of homogeneity imposed by the pixels of the increasing area. So, it is clear that results provided by such approaches, closely depend

on the selected germs. Thus we thought to use a dynamic and iterative segmentation method, selecting at the beginning of an iteration a core for each class contained in the image, allowing leading to a better partition. This core, representing of the area, doesn't consist in only one point, but includes a whole set of pixels.

4.1 Objective

The objective of this method is to generate a partition on a population whose individuals are characterized by a set of qualitative or quantitative parameters. The obtained partition must satisfy the two following conditions:

✓ The individuals of the same class must, as much as possible, look like each other.

✓ Two individuals who do not resemble each other must be in two different classes.

The population that we want to share is the image. The individuals are the pixels and the parameters which characterize each one of them are the connexity and the gray level.

4.2 Principe

The principle of this method rests on the construction of the classes, starting from NC individual subsets, constituting the NC partition cores. The individuals constituting the NC cores are either randomly predetermined or selected.

These cores allow us to define a partition of the whole of the population in the NC classes; each class includes the individuals who are closer, to its core, than to all the other cores.

From this partition we define a new set of cores which better represent the so formed classes. To this new set of cores we associate a new partition and then we start again. We stop when the formation quality does not improve any more.

In order to accelerate the classification process of this method, each core does not correspond solely to one germ, but includes a set of individuals who are the best representative ones of their classes. This property is perfectly appropriate to the image segmentation in homogeneous areas. Indeed, since the images are often noisy, a single pixel risks to not to be a good representative of its region.

Formally it is necessary to have three functions:

✓ The first calculates the distance between an individual and a core.

✓ The second function, associates to a partition of NC classes, the NC cores of Q points representative of these classes.

✓ The third function measures the quality of a partition.

Knowing these three functions, as well as the number of classes and the number of individuals by each core, the algorithm is entirely given.

There are two categories of individuals that we propose to call, similarly with the political concepts [1], elected officials and voters.

The constitution of the parties, of which the number is fixed in advance, is determined by the coalition of the voters with the groups of elected officials. Each voter is required to unite itself with one and only one group of elected officials gathered in the same party and called core. All the members of a party then have to take part in the election of their new representatives. The nævic-lesion images are composed of two parties: the lesion and the surrounding healthy skin.

4.3 Behaviour of the players

At the beginning of each negotiation period, each voter unites himself with the party which will be able to ensure the best profit to him. The profit which a pixel wants to make sure is the coexistence with the pixels which are close for him and which more resemble to him. To choose between the two parties Ai and Aj, the voter compares the profits which it hopes to obtain from each of both. But as it is difficult for him to recognize beforehand the influence of the other voters on the real profit which it will obtain when the redistribution of the profits will be carried out within its party, it estimates its profit only according to the representatives of its party.

When the elections are finished (all the members carried out their choices), the members of each party carry out the selection of their new representatives in order to maximize the total profit which will be redistributed to them. Therefore the negotiation period is finished. A new one will be necessary if the area representatives determined at the end of the last negotiation, change.

The game is thought over when the set of the voters does not change from one negotiation period to another: it means that the coalitions are unchanged even if the game lasting is prolonged. Figure 2 illustrates the followed steps.

4.4 Algorithm

Knowing that:

- Ib: is the result of the original image binairisation by one of the histogram thresholding methods previously studied in [5].
- NL and NPS: are the cores respectively representing the lesion and the healthy skin.
- "i" and "j": are counters of respectively lines and columns of the matrix image.
- p(i, j): is the pixel of coordinates "i" and "j"
- GL and GPS: are the profits calculated for the pixel "p" if it is adhered respectively to the lesion or the healthy skin.
- N and M: are respectively the number of the lines and columns of the image matrix.

The core NL is obtained by making the image Ib, undergoing "n" dilation operations. Indeed, dilation allows giving the white colour to any pixel being in the immediate vicinity of a white pixel.

Thus, by widening the white area, corresponding to the healthy skin, by "n" pixels on its entire inner boundary, we reduce the black area, corresponding to the lesion, by "n" pixels on its entire outer boundary.

The core NPS is obtained by making the image Ib, undergoing "n" erosion operations. Indeed, erosion allows giving the black colour to any pixel being in the immediate vicinity of a black pixel. Thus, by widening the black area, corresponding to the lesion, by "n" pixels on its entire inner boundary, we reduce the white area, corresponding to the healthy skin, by "n" pixels on its entire outer boundary.

The pixels that belong neither to NL nor to NPS are directly assigned to an uncertain zone. The choice of number "n" is arbitrary. But we note that more the number "n" is important, more the uncertain zone is large and thereafter the processing time will be more considerable. In this work we chose n=6.

The profit which a pixel px wants to win by uniting itself to a party P, is related to its geometrical and colorimetric resemblance with its closest pixels belonging to this party. Thus, such a profit can be modelled, for a party P, as the following expression:

$$G_{(P,px)} = \frac{1}{Rc_{(P,px)} \times (d_{(P,px)})^2}$$

Where $Rc_{(P,px)}$ is the term which gives the colorimetric resemblance between the party P and the candidate pixel px. $Rc_{(P,px)}$ is the difference between the intensity of the pixel px and the average intensity of its closest pixel vicinity belonging to the party P (cf figure 3).

$d_{(P,px)}$ is the Euclidean distance between the pixel px and the party P (cf figure 3).

5 Results

We implemented the dynamic segmentation method algorithm to evaluate it and to compare its performances with those of other techniques which we studied and tested, during a previous work. For this fact, we retained 80 images of dermatological lesions including 40 malignant and 40 benignant ones. These images were classified by clinical biopsy within the framework of a survey carried out by the CHU of ROUEN-FRANCE. Theese images were provided to us by Professor Denis of BRUCQ from the Perception - System - Information laboratory (PSI) of the University of ROUEN.

The methods of pixel classification by histogram thresholding showed their effectiveness on the nævic lesion images, because of their bimodal histograms. But since they do not take account of the space conditions of the pixels, these methods often weaken when the image presents hairs or metastases with the lesion (cf figure 4).

The methods of area growing generate an impeccable extraction of the lesions from their surrounding skin (cf figure 5). However, they are rather complex that the execution times are excessively important.

The computing time of the iterative method of dynamic segmentation, largely depends on:

- the choice of the starting classes;
- the adopted manner to choose the representative cores of the classes (called the "elected officials").

The better the choices are made, faster the convergence is. We cannot say that our choices are the best ones, but, following a consultation which we carried out near a dermatologist of the CHU of Sfax, we can affirm that the extraction reached results are amply pleasing. The results of this method are presented in figure 6 per four test images.

6 Conclusion

The dynamic segmentation method is deduced from the general concepts of the theory game which we also find in the political life. By analogy with the population individual partitioning in political coalitions, we share pixels of an image in an "a priori" known class number. We keep the concept of election which gives birth, at the end of each iteration, to a set of elected officials for each of the two

classes: lesion and healthy skin. The alliance of elected officials in one class is the set of its most representative pixels, according to a given similarity criterion. The iterations end only if the partitions do not improve according to a criterion called profit in theory game. Thus, the convergence of the algorithm depends on this criterion. As for its speed, depends on the choice of the initial classes.

The choice of the profit directly influences the results. Consequently better the profit is studied; more the segmentation can be improved.

This technique of segmentation was implemented, tested and was evaluated on 80 images of nævic lesions, including 40 malignant and 40 benignant ones. The results were suitable.

References

[1] J. Attali. Analyse économique de la vie politique. PUF, 1972.

[2] M.S. Bouhlel. Jeux, prétopologies et classification automatique. Séminaire sur les questionnaires, Université ParisVI, 1982.

[3] J.P. Cocquerez, S. Philipp, P. Bolon, J.M. Chassery, D. Demigny, C. Grafigne, A. Montanvert, R. Zéboudji, and J. Zérubia. Analyses d'images : filtrage et segmentation. Masson, Paris, 1995.

[4] M. Fontaine. Segmentation non supervisée d'images couleur par analyse de la connexité des pixels. Thèse de doctorat, Université des sciences et technologies de Lille, 2001.

[5] I. Khanfir, A. Kallel, M. S. Bouhlel, and K. Taouil. Melanoma-pattern extraction using histogram-thresholding approach. MESM03, Sharjah, Janvier 2004.

[6] B. Munier. Jeux et marché. PUF, 1973.

[7] L. Shaffarenko, M. Petrou. Automatic watershed segmentation of randomly textured color images. IEEE Trans. on Image Processing, 6(11) :1530-1543, November 1997.

[8] I. Khanfir, K. Taouil, M.S. Bouhlel, L. Kamoun Stratégie de traitement des images de lésions dermatologiques. SETIT03, Mars 2003.

[9] I. Khanfir, A. Kallel, K. Taouil, M.S. Bouhlel, L. Kamoun, "Segmentation d'images par seuillage d'histogramme: Application à l'analyse des mélanomes", JTEA, Tunisie, 2002.

[10] W. Tsai., "Moment-preserving thresholding: A new approach", Comput. Vision Graphics Image Process. 29, 1985, 377-393.

[11] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong., "A new method for gray-level picture thresholding using the entropy of the histogram", comput. Vision Graphics Image Process. 29, 1985, 273-285.

[12] T. Ridler and S. Calvard., "Picture thresholding using an iterative selection method", IEEE Trans. Systems Man Cybernet. SMC-8, 1978, 630-632.

[13] H. J. Trussel., "Comments on Picture thresholding using an iterative selection method", IEEE Trans. Systems Man Cybernet. SMC-9, 1979, 311.

[14] N. Otsu., "A treshold selection method from gray-level histogram", IEEE Trans. Systems Man Cybernet. SMC-8, 1978, 62-66.

[15] Zack, G.W., W.E. Rogers, and S.A. Latt., "Automatic Measurement of Sister Chromatid Exchange Frequency". 1977. 25(7): p. 741-753.

[16] L. Brun., "Segmentation d'images à base topologique". Thèse de doctorat, Université de Bordeaux I, 1996.

[17] A. Trémeau, N. Borel., "A region growing and merging algorithm to color segmentation". Pattern Recognition, 30 :1191–1203, 1997.

[18] A. Moghaddamzadeh, N. Bourbakis., "A fuzzy region growing approach for segmentation of color images". Pattern Recognition, 30(6):867–881, 1997.

[19] L.Shaffarenko, M. Petrou., "Automatic watershed segmentation of randomly textured color images". IEEE Trans. on Image Processing, 6(11):1530–1543, November 1997.

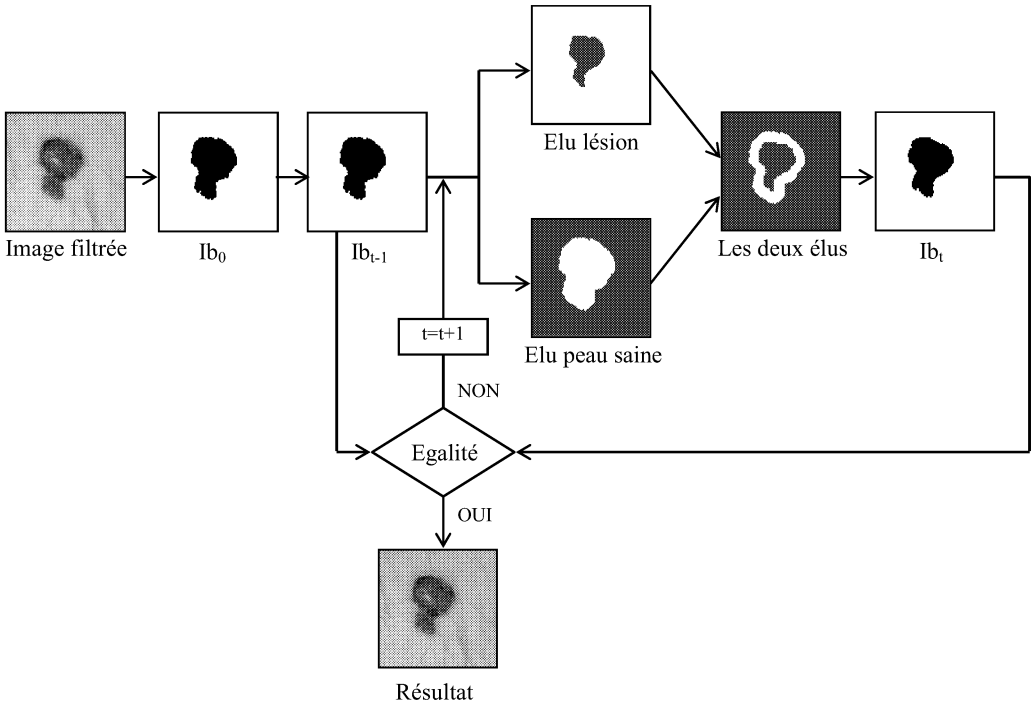


Figure 2 : Illustration of dynamic segmentation.

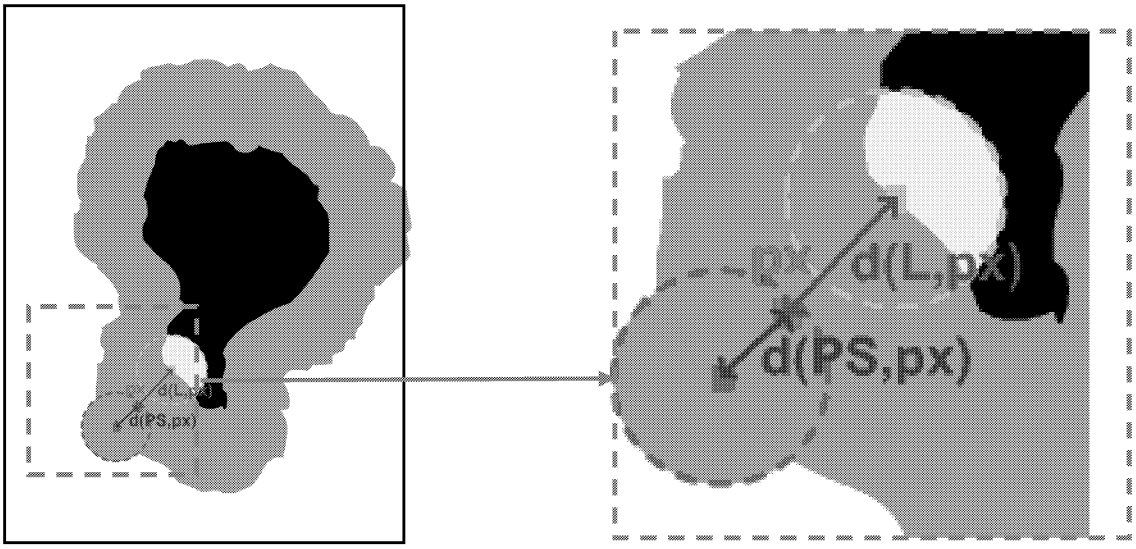


Figure 3 : Calculation of the pixel profit.

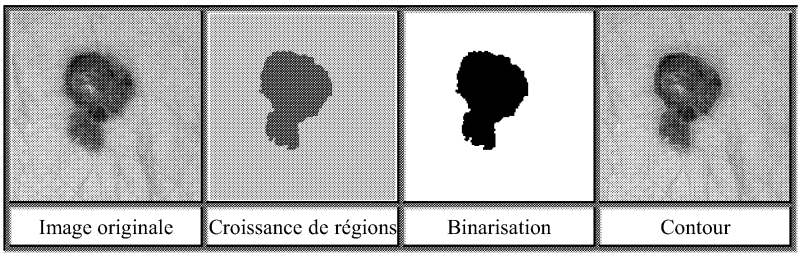


Figure 5 : Results of region growing segmentation.

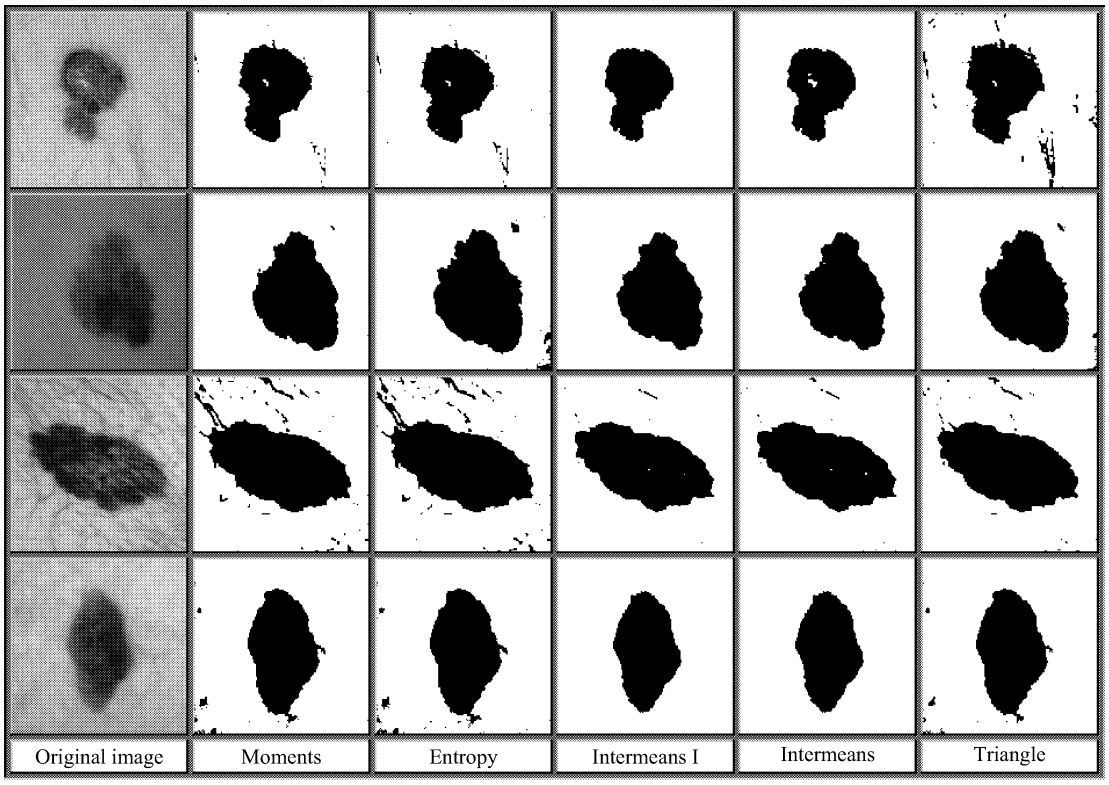


Figure 4 : Results of histogram thresholding segmentation.

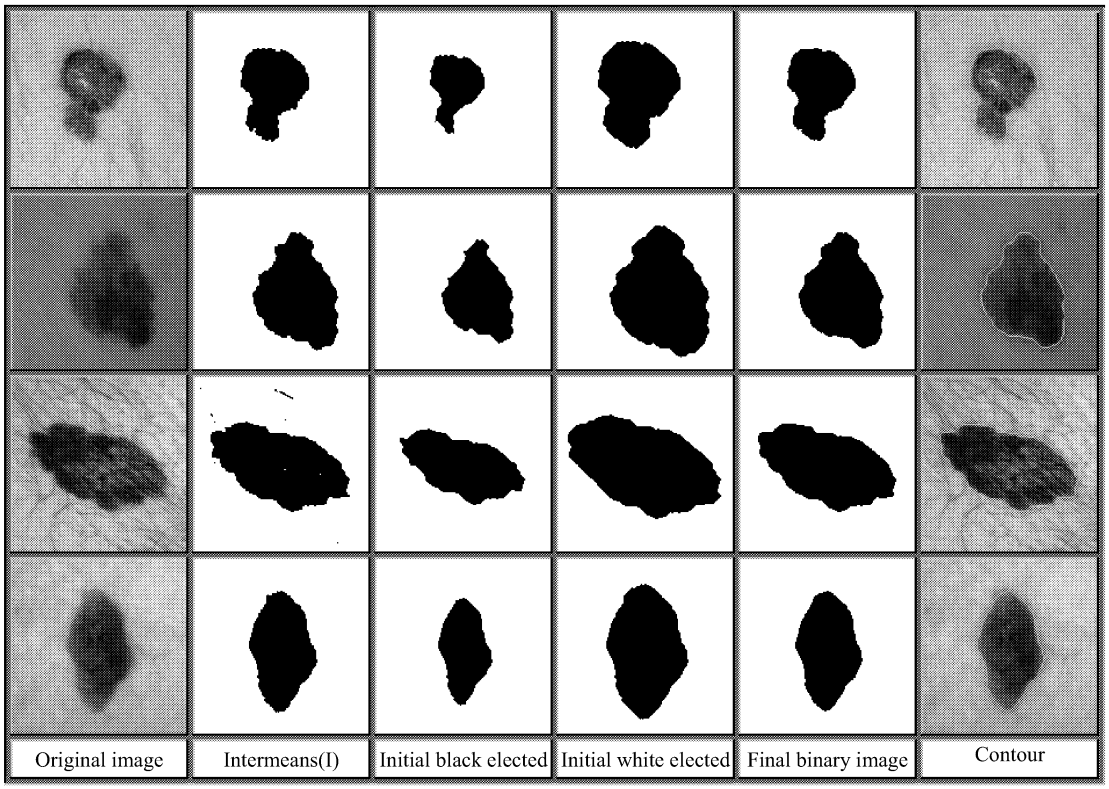


Figure 6 : Results of dynamic segmentation.

Implementation of 9/7 and 5/3 filters by the L*ifting Scheme*

Dhaha DIA
Mohamed ATRI
Rached TOURKI
Dhaha.Dia@issatso.rnu.tn Mohamed.atri@fsm.rnu.tn Rached.tourki@fsm.rnu.tn

Laboratoire d'Electronique et de Microélectronique
Faculté des Sciences de Monastir

Abstract

The preliminary work of material implemation requires the use of software. This paper studies use of *Matlab7* software of the 9/7 filter which is a transformation with loss (lossy) and 5/3 filter which is a transformation without loss (lossless) by the lifting scheme algorithm. An image of size $N \times M$ pixels is treated in two ways. First, the image is treated according only the rows of pixels. This is referred to as the one dimensional treatment of the image and it is denoted by 1D-DWT (*One Dimensional Discrete Wavelet Transform*). Then the resulting image is treated according to columns of pixels. This leads to a two dimensional treatment of the image and it is denoted by 2D-DWT [1]. The lifting scheme (LS) being an algorithm with the perfect reconstruction property, a threshold is applied to the coefficients of the filters in order to compute the *PSRN* (Peak *Signal- to-Noise Ratio*). Graphs of *PSRN* and entropy as a function of the threshold are provided for the two types of transformation of both filters. This article is divided in five parts. First, a brief introduction to the LS is presented. Second, we develop some theory results of the LS. Third, we present our method of implementation. Forth, results and discussions are given. Finally, a conclusion is provided.

Key words: *lifting scheme*, 9/7 filter, 5/3 filter, DWT, Matlab7

INTRODUCTION

High demand on the quality of transmission, the rate, the coots via internet or satellites of images among others required by an eventual implementation on *FPGA* or *ASCI* impose many constraints. In order to go around some of these constraints preliminary software treatments are required. The convolution method has long been the classical tool of image filtering by DWT thanks to filter bancs [6] algorithm which necessitates a large number of algebraic calculations. The LS created in 1994 by Win Sweldens allows reducing the number of calculations by bypassing the dilation and the translation properties of the wavelets. Therefore, the LS can be thought of as a simple "classical wavelet" with steps of predictions and up dating [3]. The LS being an algorithm of a material implementation, it is subject in this study to a pre treatment using matlab7. The steps of factorization in polyphase (1), of predictions and up dating are essential in the LS algorithm. The numbers of predictions and up dating vary according to the filter considered, for example, 9/7 filter require two predictions and two up dating (fig.2) while 5/3 filter require only one predictions and one up dating. The factorization in polyphase requires estimation of coefficients of the steps of predictions and up dating. Therefore, two types of coefficients estimates are used: the rational and irrational estimates [4].

LIFTING SCHEME

The lifting scheme is an algorithm destined at implantation material hardware of *Discrete Wavelet Transform* (DWT) [5]. It is constituted of steps of predictions and up dating describe by the fig.1. The inverse transform is obtained, to go run to right at left, which inverse the K , $\frac{1}{K}$ and to change the sign [2].

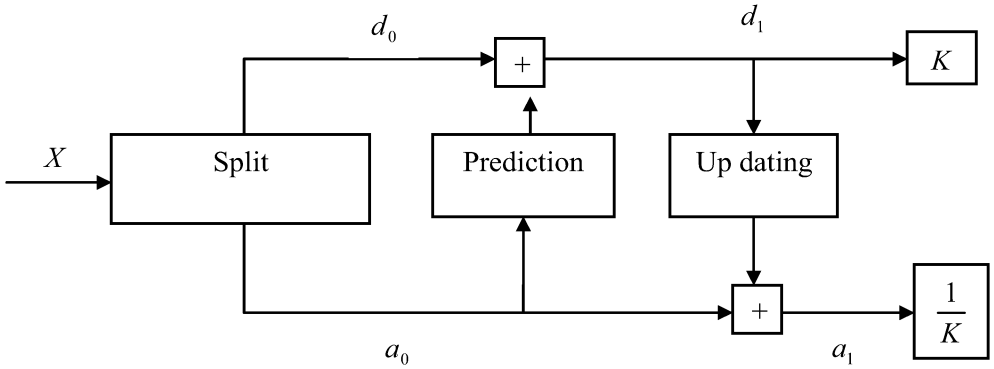


FIGURE 1: The forward transform of *lifting scheme*

The polyphase representation, of discrete-time filter $H(z)$ is defined as:

$$H(z) = H_e(z^2) + z^{-1}H_o(z^2) \quad (1)$$

where $H_e(z)$ and $H_o(z)$ are respectively obtained from the even and odd coefficients of $h[n] = Z^{-1}\{H(z)\}$, where Z denotes the zeta transform. Expressing the synthesis filter $H(z)$ (low pass) and $G(z)$ (high pass) through their polyphase representation we obtain the polyphase matrix:

$$P(z) = \begin{bmatrix} H_e(z) & G_e(z) \\ H_o(z) & G_o(z) \end{bmatrix} \quad (2)$$

($\overset{*}{(}P(z)$ can be analogously defined for the analysis filters). The filters $H_e(z), H_o(z), G_e(z)$ and $G_o(z)$ are Laurent polynomials. As the set of all Laurent polynomials exhibits a commutative ring structure, within which polynomial division with remainder is possible, long division between two Laurent polynomials is not unique operation. To decompose $P(z)$ the Euclidean algorithm can be used to obtain:

$$P(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & \frac{1}{K} \end{bmatrix} \quad (3)$$

where $s_i(z)$ and $t_i(z)$ are respectively primary lifting steps and dual lifting steps filters and K is a constant. In fig.2 we have four steps: two primary lifting steps (predictions) and two others dual lifting steps (up dating) of 9/7 filter [7].

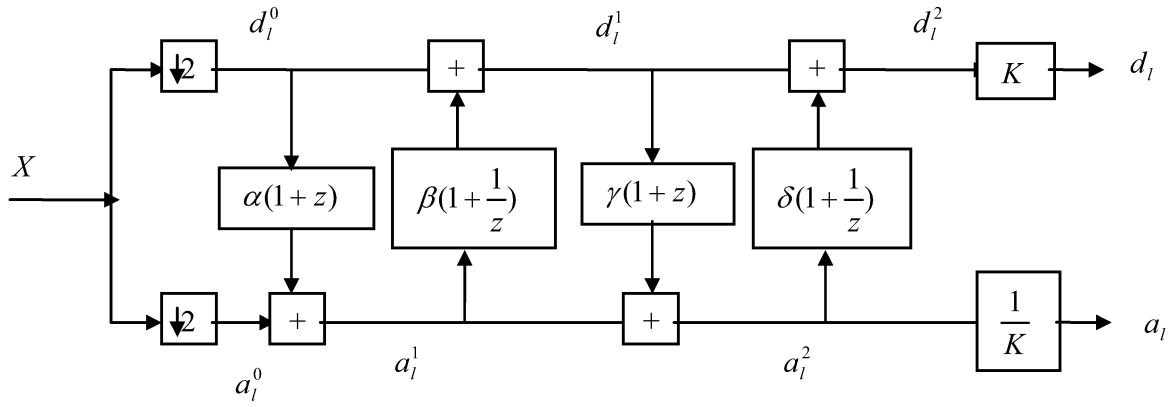


FIGURE 2: Factorisation of bi orthogonal wavelet 9/7 by *lifting scheme*

METHOD OF IMPLEMENTATION

The two types of wavelet transformation are treated in this study: 1D-DWT and 2D-DWT according rows and columns of image. The image is processed as a matrix in Matlab. The image rows are treated successively when a 1D-DWT transformation is achieved. Every row is considered as a one dimensional signal. It is splitted in two parts, containing respectively even and odd samples (fig.1); calculated according to equations (4 and 5) [3]:

For the odd samples called details:

$$d_k = x_{2k+1} - \frac{(x_{2k} + x_{2k+2})}{2} \qquad (4)$$

For the even samples named approximation:

$$a_k = x_{2k} + \frac{d_{k-1} + d_k}{4} \qquad (5)$$

Finally we obtained by the forward transform:

$$G_1 = K \times d_k \qquad (6)$$

$$G_2 = \frac{a_k}{K} \qquad (7)$$

where $k \in Z$ and x is the original signal. The equations (1) and (2) are named respectively, the prediction, the up dating and the last ones are normalised odd and even coefficients by K and $\frac{1}{K}$.

The 2D-DWT is the double transformation of 1D-DWT by rows and columns (fig.3).

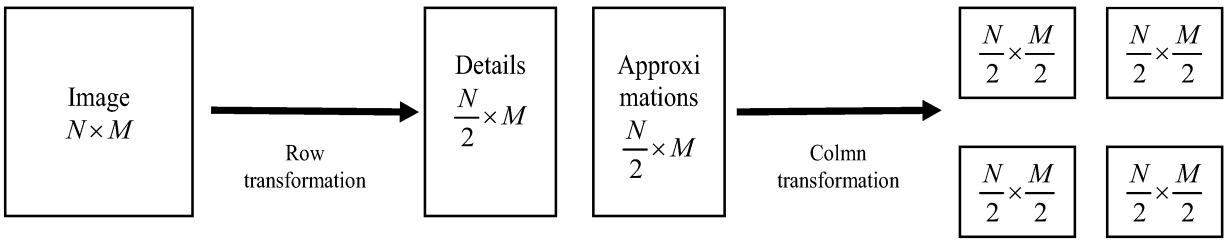


FIGURE 3: Forward bloc of 2D-DWT with *lifting scheme*

The quality of different images was accompanied the *PSRN* calculated using the expression for an image of 256 Gray level is:

$$p = 10\log_{10}\left(\frac{255^2}{\frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}(x(i,j)-y(i,j))}\right) \quad (8)$$

The compression rate expression is:

$$T_c = \frac{m}{m-n} \quad (9)$$

which m is the pixel total number of image and n the number zero pixels of the associated image.

The entropy expression is:

$$H = \sum_{i=0}^{T} p(n_i)\log 2(\frac{1}{p(x_i)}) \quad (10)$$

witch T is the total number of Gray level and $p(n_i)$ is the probability of number Gray level n_i.

RESULTATS AND DISCUSSIONS

In this section we give some results of 1D-DWT and 2D-DWT transformation for six different images of 255 Gray level (see tables). It is accompanied by statistical calcul like, the *PSRN* and the zero number indicating when the threshold was applied to different image coefficients.
The graphs of the *PSRN* and entropy vary with thresholds, for the two type transformations of two filters as shown in next figures.

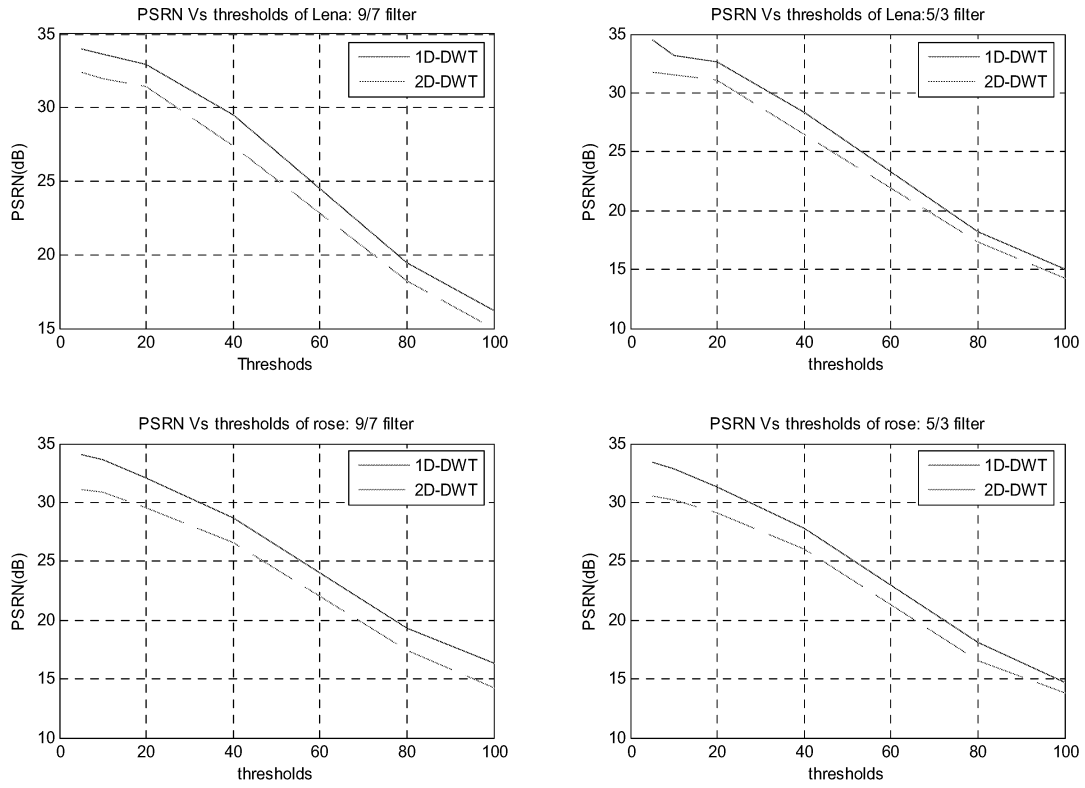


FIGURE 4: The graphs *PSRN* function of threshold according the two transformations: In up the Lena image and down the rose image, left 9/7 filter and right 5/3 filter.

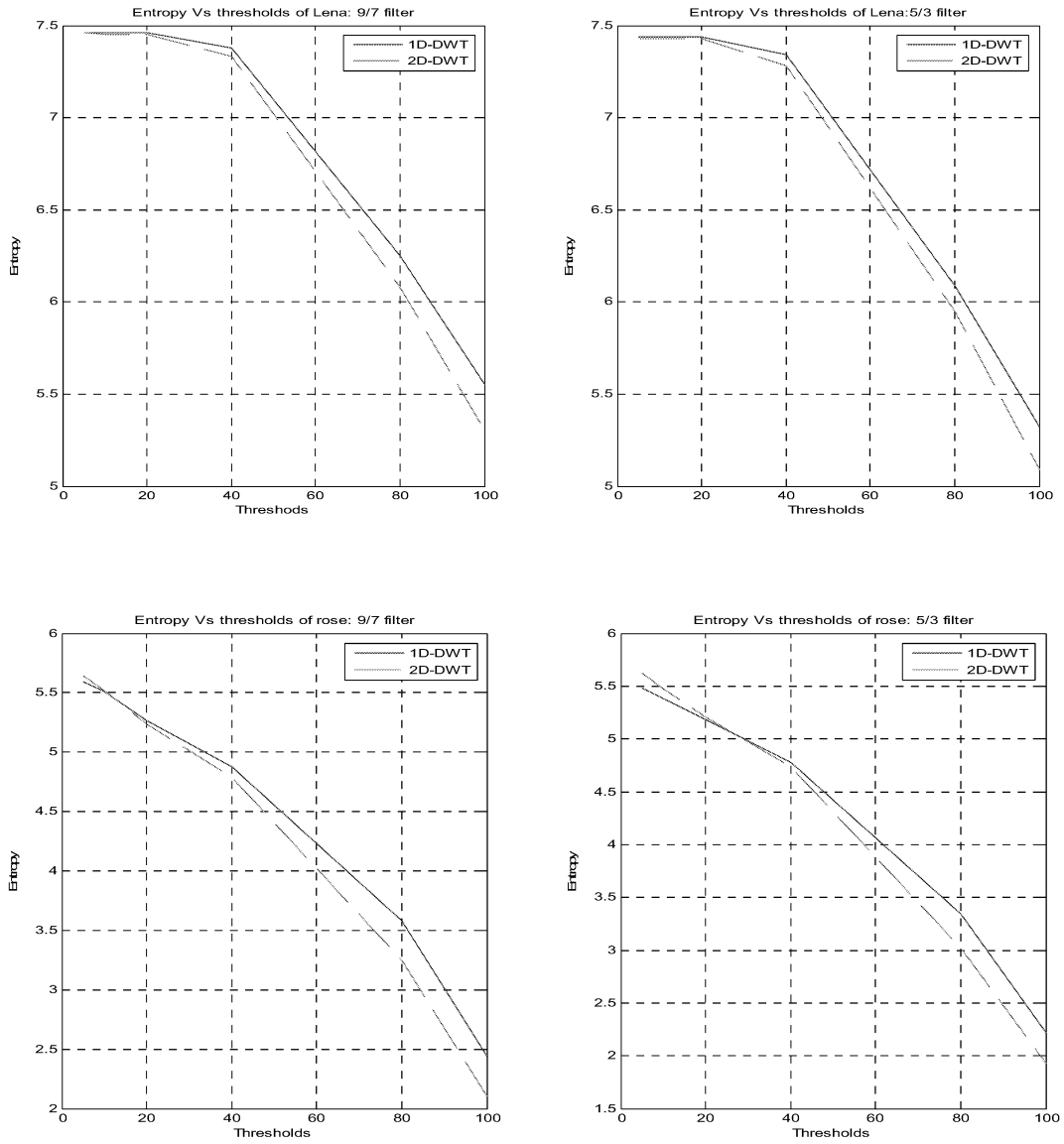


FIGURE 5: The graphs entropy function of threshold according the two transformations: In up the Lena image and down the rose image, left 9/7 filter and right 5/3 filter.

After the results of these transformations, the 1D-DWT present the quality image synthesis better than 2D-DWT, in opposite about the compression rate (see zero number in tables). The LS is destined at implantation hardware, contrary at compression, this confirmed by the results of tables in down. In table II, the 5/3 filter present the *PSRN* rate infinite, identity equal the synthesis image is equivalent at the original image, conformed the propriety of lossless compression, in opposite propriety (lossy) at the 9/7 filter (table I). In compared the threshold (figure 4) Lena which rose, the results show, when the threshold is less 20 Gray level, the quality of the both image was degraded for the two filters. The image rose, present the rate compression more than Lena, at different threshold of two filters. In the

graphs entropy function of threshold of the two images Lena and rose, varying differently. The entropy of rose is less than Lena; it is between to 2 at 6 and Lena it show to 5 at 7,5 function of threshold. The entropy is constant when the threshold is less 40 level Gray, and it decrease quickly, on the other hand the rose image, his graph entropy in function of threshold decrease of the two filters (figure 5) for the types transformations.

CONCLUSION

This study provides a conformation that the 9/7 and 5/3 filters are transformation with loss and transformation without loss, respectively. In addition, it shows that the LS are not an ideal algorithm of image compression. This work is based on a classical method of image treatment using two dimensions. It gives us preliminary information on an eventual material implantation of the two types of filters. Various tests of estimates of the coefficients of both filters led to choice of rational coefficient estimates thanks to a better quality of their synthesized images.

REFERENCES

[1] S.Barua, J.E. Carletta, K.A. Kotteri, A.E.Bell " An efficient architecture for lifting-based two-dimensional discrete wavelet transforms" Integration, the journal 38 (2005) 341-352 21 July 2004.

[2] Kishore Andra, Chaitali Chakrabarti, Member, IEEE and Tinku Acharya, Senior Member, IEEE "A VLSI Architecture for Lifting-Based Forward and Inverse Wavelet Transform" IEEE Transactions on signal processing, vol.50, NO.4, APRIL 2002

[3] Camille Diou "Contribution à l'intégration du silicium de la transformée en ondelette; en vu d'une implantation matériel" Université Montpellier II, 12 Décembre 2000.

[4] S.Barua, K.A.Kotteri, A.E.Bell and J.E.Carletta"Optimal Quantized Lifting Coefficients for the Wavelet" 0-7803-8484-9/04/$20.00 2004 IEEE.

[5] Marco Grangetto, Enrico Magli, Gabriella Olma "Finite Precision wavelets for image coding: Lossy and Lossless, compression performance evaluation" Dipartimento di Electronica-Politecnio di Torino, Corso Duca degli Abruzzi 24-10129 Torino-Italy.

[6] Kishore A. Kotteri, *Student Member*, *IEEE,* Amy E. Bell, *Member, IEEE*, and Joan E. Carletta, *Member, IEEE* " Design of Multiplierless, High-Performance, Wavelet Filter Banks With Image Compression Applications" IEEE Transactions on circuits and systems-I: Regular Papers, vol.51, No.3, March 2004.

[7] M. Martina, G. Masera, G. Piccinini, M. Zamboni « A VLSI architecture for IWT (Integer Wavelet Transform) » Dipartimento di Electronica-Politecnio di Torino, Corso DUca degli Abruzzi 24-10129 Torino-Italy.

Table I: Results transformation of 1D-DWT and 2D-DWT by *lifting scheme* of 9/**7 filter** in six images.

Images	1D-DWT			2D-DWT		
	Pixels Number	Zero Number (%)	*PSRN* (dB)	Pixels Number	Zero Number (%)	*PSRN* (dB)
Lena	65536	0	47,74	65536	0	45,36
Clown	64000	0	49,12	64000	0	45
Guépard	64000	0	47,88	64000	0	42,53
Mouse	64000	0,6	60,65	64000	0	55,81
Rose	64000	28,47	47,92	64000	21,19	44,41
Lisa	64000	0	59,73	64000	0	56,27

Table II: Results transformation of 1D-DWT and 2D-DWT by *lifting scheme* of **5/3 filter** in six images.

Images	1D-DWT			2D-DWT		
	Pixels Number	Zero Number (%)	*PSRN* (dB)	Pixels Number	Zero Number (%)	*PSRN* (dB)
Lena	65536	0	∞	65536	0	∞
Clown	64000	0	∞	64000	0	∞
Guépard	64000	0	∞	64000	0	∞
Mouse	64000	0,88	∞	64000	0,88	∞
Rose	64000	36,59	∞	64000	36,59	∞
Lisa	64000	0	∞	64000	0	∞

Table III: Results of threshold of Lena, when the total number pixels, is 65536, at two type transformation of two filters.

Thresholds (In Gray level)	**9/7 Filter**				**5/3 Filter**			
	1D-DWT		2D-DWT		1D-DWT		2D-DWT	
	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)
5	0	34,03	0	32,40	0	34,45	0	31,80
10	0	33,65	0	32	0	33,15	0	31,55
20	0	32,95	0	31,39	0	32,58	0	31,04
40	2	29,50	3,39	27,40	3,09	28,34	4,63	26,47
80	21,68	19,41	24,34	18,22	24,26	18,24	26,44	17,35
100	31,54	16,15	35,51	15	35,26	15,04	38,73	14,20

Table IV: Results of threshold of rose, when the total number pixels, is 65536 at two type transformations of two filters.

Thresholds (In Gray level)	**9/7 Filter**				**5/3 Filter**			
	1D-DWT		2D-DWT		1D-DWT		2D-DWT	
	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)	Zero Number (%)	*PSRN* (dB)
5	34,48	34,10	36,11	31,33	36,67	33,35	36,07	30,58
10	37,65	33,57	37,68	30,87	38,43	32,81	40	30,24
20	40,56	32,05	41	29,60	41,20	31,28	41,18	29,10
40	44,96	28,70	46,04	26,54	45,76	27,78	46,48	26,05
80	60,46	19,30	65,38	17,40	63,56	18,10	68 ,27	16,60
100	74,78	16,37	79,41	14,20	77,43	14,70	81,27	13,77

FACE RECOGNITION USING LINEAR ANALYSIS TECHNIQUES AND NEURAL NETWORKS BASED APPROACHES

Ichraq Ben Aouicha, Tarek Tray and Dorra Sellami Masmoudi

Research Unit on Intelligent Control, Design and Optimization of Complex Systems
Ecole Nationale d'Ingénieurs de Sfax – Tunisia

dorra.masmoudi@enis.rnu.tn

Abstract:
In this paper, we are concerned with face recognition techniques using principal component analysis (PCA), Fisher linear discriminant Analysis (FLDA) as linear approaches and Neural Networks. Known as one of the efficient techniques used in dimensionality reduction, PCA is based on information theory concepts, seeking a computational model that best describes a face, by extracting the most relevant information contained in that face. FLDA is a classification method that projects high-dimensional data onto a reduced space and performs classification in it. The projection maximizes the distance between the means of all classes while minimizing the variance within each class. Although, offering a good reduction in dimensionality, the reduced dimension remains too high for a decision making process. Neural Networks represents a good solution for decision making by achieving numeric classification. It is shown in this paper that hybrid solution associating dimension reduction techniques and Neural Networks reduces computational efforts required in Neural Networks image classifiers. Indeed, although giving satisfactory solutions, Neural Networks based approaches require high computations in face recognition classification problems.

1. INTRODUCTION

Face recognition can be applied for a wide variety of problems like image and film processing, human-computer interaction, criminal identification …etc. This has motivated researchers to develop computational models to identify the faces, which are relatively simple and easy to implement.

A variety of approaches are used in face recognition. Having different degrees of complexity, good or less efficiencies are reported.

Nearest-Neighbour- Template Matching is the most naïve method of face recognition. Individuals are represented by a set of complete images of their face. An input face is identified through comparisons with each image in the database. Finding the image for which the template matching gives the highest response identifies the person, given that it is significant enough to represent a match [1] [2]. Many problems arise using this method:

- Template Matching has an inability to deal with changes in intrinsic or extrinsic factors. A large search space would be required to handle such transforms. Images of all possible factors would have to be included. Not only would this be time consuming to initially produce the database but also computations on this huge search space would be far too slow.
- When using this method it is necessary to look at every pixel in each of the images, taking each pixel to be a point in the image space. With an image of size 192x128 there

are 24,576 pixels. The speed of computations on data this size will be far too slow, especially with a large database.

- As discussed in first, point to get good results with this method it is necessary to use a huge database. This would require a large amount of memory and is not efficient.

Feature-based face recognition uses methods that locate pre-defined distinctive features of faces, such as eyebrow thickness and their vertical position or nose position and width. The corresponding figures are stored as a feature vector. The vector is compared with those of the known database to find the closest match. As with nearest-neighbour template matching this method also has limitations which bring doubt to its accuracy. The most important of which is that the feature detectors are unable to cope with facial occlusions. Standard feature-based methods require far less memory and detection is completed at much higher speeds compared to template matching. But template matching has been shown to produce much higher detection rates. Using efficient and data independent feature detectors such as Principal Components Analysis, it will be shown that there is potential for even higher detection rates [1].

Accordingly, Neural Networks approaches are the best candidates to meet the requirements for imitating the Human distinguished qualities regarding face recognition capabilities. Such solution can give satisfactory results but its implementation is cumbersome. Linear techniques such as PCA and FLD can help undergoing these limitations by offering high dimensionality reduction.

We deal in this paper with linear techniques applied for features extraction and dimentionality reduction. Such techniques such as PCA and FLDA [3] [4]can be associated with Neural Networks to help decision making.

2. OVERVIEW ON CONVENTIONAL FACE RECOGNITION METHODS

2.1. Eigenfaces method

It is based on an information theory approach that decomposes face images into a small set of characteristic feature images called “eigenfaces”, which are actually the principal components of the initial training set of face images [3]. Recognition is performed by projecting a new image into the subspace spanned by the “eigenfaces” (“face space”) and then classifying the face by comparing its position in the face space with the positions of the known individuals [4].

Each face image is considered as a two-dimensional mxm array of intensity values. It is then converted to a vector of dimension m^2.

$$X_i = \begin{pmatrix} X_{i,1} \\ X_{i,2} \\ . \\ . \\ X_{i,m^2} \end{pmatrix}$$

The training set of n face images is denoted by:

$$X = [X_1 X_2 X_3 ... X_n] .$$

The covariance matrix is defined as follows:

$$C = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^T$$

Where $\overline{X}$ is the mean image:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Then, the eigenvalues and eigenvectors of the covariance C are calculated. The new subspace w_{PCA} contains the eigenvectors corresponding to the r largest eigenvalues.

$$w_{PCA} = \begin{bmatrix} u_1^T \\ u_2^T \\ .. \\ .. \\ u_r^T \end{bmatrix}$$

For the set of images X, the corresponding eigenspace is obtained by the projection:

$$y = w_{PCA}.X$$

2.2. Fisherfaces method

Actually, PCA shows a great performance in dimensionality reduction but doesn't provide any information for class discrimination. So the question is: how do we utilize the label information in finding informative projections. Accordingly, Fisher-LDA considers maximizing the following objective [6] [7]:

$$J(w_{FLD}) = \frac{w_{FLD}^T S_B w_{FLD}}{w_{FLD}^T S_w w_{FLD}}$$

Where S_B the "between classes scatter matrix" and S_w is the "within classes scatter matrix" defined as follows:

$$S_B = \sum_c n_c (\mu_c - \overline{X})(\mu_c - \overline{X})^T$$

$$S_w = \sum_c \sum_{i \in c} (X_i - \mu_c)(X_i - \mu_c)^T$$

Where:

$$\mu_c = \frac{1}{n_c} \sum_{i \in c} X_i$$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_c n_c \mu_c$$

The optimal subspace is obtained by calculating the generalised eigenvectors:

$$S_B w_{FLD} = \lambda S_w w_{FLD}$$

$$\Rightarrow$$

$$S_w^{-1} S_B w_{FLD} = \lambda w_{FLD}$$

Thus, the feature vectors for any face images X in the most discriminant sense can be calculated as follows:

$$P = w_{FLD} X$$

3. NEURAL NETWORK BASED APPROACH

Neural networks-based approaches are commonly used in face recognition. It consists on finding the relevant characteristics of face images. The learned characteristics, in the form of discriminant functions (i.e. non-linear decision surfaces), are subsequently used for face recognition. Conventionally, face images are projected to a low-dimensional feature space and nonlinear decision surface is formed using multilayer neural networks for classifications and recognition. Neural networks have also been used successfully for face recognition problem [7]. The advantage of using the neural networks for face recognition is that the networks can be trained to capture more knowledge about the variation of face patterns, and thereby achieving good generalization. The main drawback of this technique is that the networks have to be extensively tuned to get exceptional performance. Among the neural networks approaches for face recognition, multilayer perceptron (MLP) with back propagation (BP) algorithm has been mostly used.

4. EXPERIMENTAL RESULTS

Our experiments were performed on the face database which contains a set of face images taken between April 1992 and April

1994 at the Olivetti Research Laboratory (ORL) in Cambridge University, U.K. There are 400 images of 40 individuals. For some subjects, the images were taken at different times, which contain quite a high degree of variability in lighting, facial expression (open/closed eyes, smiling/no smiling… etc.), pose (upright, frontal position etc), and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to 20° [7].



Figure 1: Various view images of one person in Olivetti database.

The experimental results presented in this section are divided into two parts. The first one evaluates the effectiveness of conventional methods applied directly. The second part reports the performance of the use of neural network.

4.1. Conventional methods results

In the following experiments, a total of 200 images were randomly selected as the training set and another 200 images as the testing set, in which each person has five images.

- Eigenfaces approach

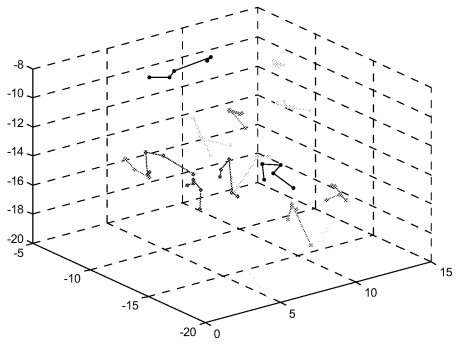


Figure 2: Projection on the first components

4.2. Fisherfaces approach results

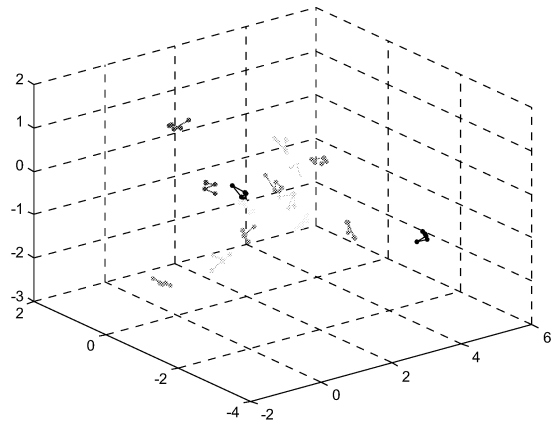


Figure 3: Projection on the first components

4.3. PCA & FLD results

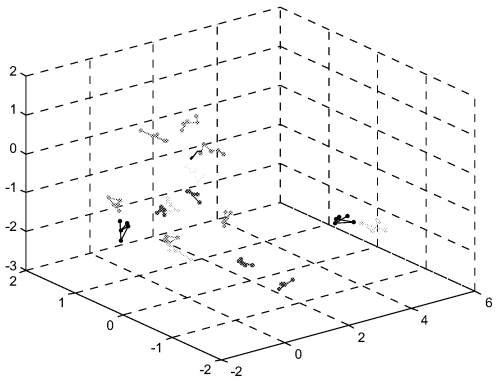


Figure 4: Projection on the first components

A 3D plot has been produced to help showing discrimination. The plot presents the distribution of images randomly selected. It is necessary to note the difficulty to take decision with such projection. To solve the problem, Multi Layer Perceptron (MLP) is used for numeric classification.

4.4. Neural network

One hidden layer neural network is used in this application with the following parameters:

- Features vectors correspond to the input layer.
- Classes correspond to the output layer
- Transfer function used is sigmoid.

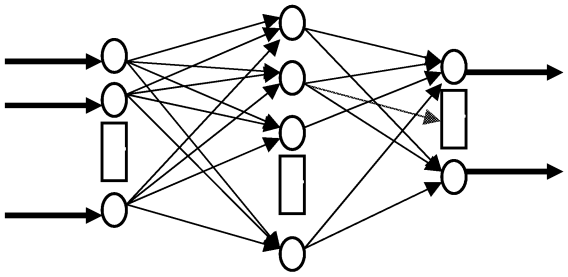


Figure 5: Multi layer perceptron

The back propagating MLP cycle is as follows:

- Calculate the network outputs.
- Compare the network outputs with the target function.
- Calculate errors.
- Propagate the errors backward through the network.
- Adjust the connecting weights according to the error information.

The training cycle is repeated until finding minimum error.

Classes	Hidden cells	Learning time	Error on test images
6	7	21mn 13 s.	5.77%
8	9	33mn 12 s.	6.76%

Table 1: Simulation with MLP

The previous algorithm gave good results for a small training set, nevertheless it was inefficient for a large data. As a solution, we applied dimensionality reduction techniques.

4.5.PCA & MLP results

Applying PCA as dimensionality-reduction technique provides a generalization of results for a large data with a low discrimination rate. In choosing the projection which maximizes the total scatter, the PCA retains unwanted variations caused by lighting, facial expression, and other factors. Accordingly, the features produced are not necessarily good for discrimination among classes

Thus, applying FLD looks to be essential.

Classes	Hidden cells	Learning time	Error on test images
4	**5**	**1.75s.**	**21.54%**
4	6	1.67s.	23.8%
6	**9**	**2.13 s.**	**16.16%**
6	11	1.44 s.	17.9 %
10	**12**	**26.6¨s.**	**9.84%**
10	15	6.6 s.	12.84%
20	25	11mn 36 s.	10.28%
20	**30**	**13mn 43 s.**	**9.86%**

Table 2: Simulation with PCA & MLP

4.6. PCA & FLD & MLP results

Classes	Hidden cells	Learning time	Error on test images
5	7	3.8 s.	9.38%
5	**8**	**1.8 s.**	**6.2e-6%**
10	12	21.5 s.	4.47%
10	**15**	**4.531 s.**	**3.43e-6%**
10	17	24.406 s.	4.47%
20	**23**	**1mn 25 s.**	**2.36e-6%**
20	25	5mn 20 s..	2.24%
25	28	14mn 17 s.	1.79%
25	30	5mn 49 s.	1.79%
25	**35**	**29 mn 10 s.**	**2.47e-6%**
30	35	42mn 3s.	1.49%
30	38	1h 5mn 24 s.	1.49%

Table 3: Simulation with PCA & FLD & MLP

Applying PCA & FLD & MLP technique, we get a generalization which is more important with a very low error rate.

4. CONCLUSION

In this paper, a general view and application of face recognition techniques are presented. Approaches used are divided into two categories. The first one is based on principal components analysis and Fisher's linear discriminant. We have to note that it is difficult to take decision with those conventional methods. Applying neural networks provides a numeric classification for data. The principal advantage of the method is the high speed rate of identification-time, whereas the learning time is growing up with the number of classes.

6. REFERENCES

[1] Face Recognition Techniques and the Implications of Facial Transformations "David Julian Oziem" University of Bristol May 2002.

[2] http://www.techno-science.Net /forum/ viewtopic.php?t=820

[3] http://mercier.homelinux.net/~ten0k/ DEA/biblio_autre/eigen/mturk-VPR91.pdf

[4] http://www.hds.utc.fr/~fdavoine/GDR visages/colineau.pdf

[5]http://www.cs.njit.edu/~liu/papers/icpr98 .pdf

[6]http://www.cedar.buffalo.edu/~srihari/C SE555/Chap3.Part4_FisherLD.pdf

[7] Face Recognition with Radial Basis Function (RBF) Neural Networks. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 3, MAY 2002.

DATA MODELLING AND SIMULATION

THE CAPABILITY OF REDUCING THE PAPR BY TURBO CODING

Omar Daoud,
Marwan Al-Akaidi and
Jon Ivins
School of Engineering & Technology, De Montfort University, Leicester, UK
omard@dmu.ac.uk,
mma@dmu.ac.uk and
jri@dmu.ac.uk.

KEYWORDS
Beyond 3G, Orthogonal Frequency Division Multiplexing, Peak-to-Average Power Ratio and Turbo Coding.

ABSTRACT

This paper presents a new algorithm for reduction the Peak-to-Average power Ratio (PAPR) through the joint use of turbo coding and choosing the combination of the bits that gives the best reduction for the PAPR. Indeed, the new technique shows that the PAPR can be reduced in spite of the changing of different system parameters, such as, the mapping techniques and the MIMO encoders.

INTRODUCTION

Due to the rapid growing of demand for mobile communication and computing, there is increased demand for Internet access. The key to achieving this increase is the development and deployment of high performance radio systems that support high data rates with wide area coverage. Many problems appear in conventional single-carrier communication systems for high data rates due to short bit duration. Therefore, in high data rate systems the channel affects increase, which gives an advantage to the robust systems against the multipath fading, such as Orthogonal Frequency Division Multiplexing (OFDM). Moreover, OFDM has become a popular technique to support high-speed wireless data for various systems. In such a system, the deleterious effect of fading is spread over many bits; therefore, instead of a few adjacent bits completely destroyed by the fading, it is more likely that several bits are only be slightly affected by the channel. Furthermore, when the data is transmitted at high bit rates over mobile radio channels there exists Inter Symbol Interference (ISI) due to the channel impulse response extending over several symbol durations. The OFDM can be gainfully employed to mitigate the effects of ISI by attaching a part from the symbol to be at the beginning (Stuber G. 2001; Nee R. and Prasad R., 2000).

To gain additional advantages from the use of OFDM, such as increasing the diversity gain, and capacity, OFDM modulation techniques may be combined with multiple antennas at both ends, as shown in Figure 1. In OFDM systems, the transmitted signal occasionally exhibits very high peaks. The principle disadvantage is that the peak-transmitted power may be up to N times the average power, with N, the number of subcarriers. A large PAPR introduces an increased complexity of the ADC, DAC, and degrades the efficiency of RF power amplifier. Linear amplifiers that can handle the peak power are less efficient. Hard limiting of the transmitted signal generates intermodulation among the subcarriers and results in out-of-band radiation (Cimini L. and Sollenberger N. 2000).

Various techniques have been proposed in the literature to optimise the PAPR problem in the OFDM system, such as coding techniques, weighting functions, partial transmission sequence, clipping, windowing and selected mapping. In this paper, we consider a new usage of the coding technique to accomplish the optimisation process for the PAPR problem (S. Shepherd, et al. 1998; Nee R. and Prasad R., 2000). Finally, through numerical investigation using computer simulation offer a considerable reduction in PAPR.

THE SYSTEM BLOCK DIAGRAM

The complete block diagram of a physical-layer-based forward section is shown in Figure 1. The data source, assumed to be a sequence of discrete data coming from the upper layers, goes through a clustering technique to be clustered in to *M* sequences, where *M* is the number of clusters.

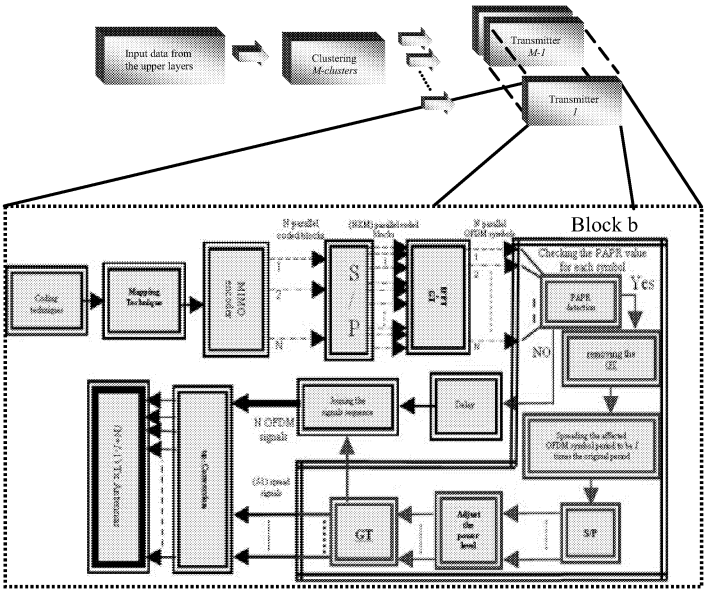


Figure 1: The forward section block diagram.

The data in each cluster will be $\frac{1}{M}$ times the total bit rate:

$$\{d_0,d_1,d_2,\ldots.\} \qquad (1)$$

The output of the encoder is fed to a mapping technique, the input data arriving at the rate R_1 will be in the P parallel information symbols. Therefore, the number of bits in each P sequence will depend on the constellation technique. For example, in 16-QAM modulation technique they will be 4 bits of information for each symbol, so the rate will be reduced to (R_1/4) bits/sec. That means the output will be a vector of P complex numbers given by:

$$\{ c_0,c_1,\ldots., c_{P-1}\} \qquad (2)$$

The data will go through a MIMO encoding techniques to achieve the benefit that we want from using the MIMO technique, either to increase the diversity gain or to increase the system capacity. After that, the data is fed into the IFFT with P-points. The output of the IFFT will be given by:

$$\{ f_0,f_1,\ldots., f_{P-1}\}$$

where:

$$f_i = a_i + b_i = \sum_{v=0}^{P-1} c_v e^{j\frac{2\pi}{P}vi} \; ; \; i= 0,1,\ldots,P\text{-}1 \qquad (3)$$

At this stage the OFDM symbol will be generated, and as we said before, that the OFDM signal occasionally exhibits very high power peaks. We will optimise the PAPR before the signal goes to the up-conversion stage. The method that is used to optimise the PAPR will be discussed in the next section.

FORMULATION OF PAPR PROBLEM

The transmitted OFDM signal may be modelled as:

$$f_i = a_i + b_i = \sum_{v=0}^{P-1} c_v e^{j\frac{2\pi}{P}vi} \; ; \; i= 0,1,\ldots,P\text{-}1 \qquad (4)$$

and it can be written as:

$$f_i(t)= \sum_{v=0}^{P-1}(A\cos(2\pi f_v t) - B\sin(2\pi f_v t)) + j\sum_{v=0}^{P-1}(A\cos(2\pi f_v t) + B\sin(2\pi f_v t)) \qquad (5)$$

The PAPR of the continuous modulated symbol in (5) can be written as:

$$\text{PAPR}= \frac{\max\{|f_i(t)|^2\}}{E\{|f_i(t)|^2\}} \qquad (6)$$

From this equation, we can find a relationship between the PAPR and OFDM symbol period by calculating the average power of the inverse fast Fourier transformed signal for a digitally modulated data as follows

$$\text{avg}(f(t)) = E\{|f_i(t)|^2\} = \frac{1}{T}\int_0^T |f(t)|^2 dt \qquad (7)$$

Substituting f(t) from (4) in (7)

$$\text{avg}(S(t))= \frac{1}{T}\int_0^T \left|\sum_{v=0}^{P-1} c_v * \exp(j*2*pi*\frac{v}{PT}\ t_v)\right|^2 dt$$

$$= \frac{1}{T}\int_0^T (\sum_{v=0}^{P-1} c_v * \cos(2*pi*\frac{v}{PT}\ t_v))^2 + j(\sum_{v=0}^{P-1} c_v * \sin(2*pi*\frac{v}{PT}\ t_v))^2 dt$$

$$= \frac{1}{T}\int_0^T (\sum_{v=0}^{P-1} c_v^2) dt$$

Consider the used modulation is a BPSK with $|c_v| = 1$, when t $\in$[0, T]. This leads to direct relationship between the average power and the total number of subcarriers N can be taken directly from

$$\text{avg}(f(t)) = P\ \frac{1}{T}\int_0^T c_v^2 dt = P \qquad (8)$$

The achieved result from (8) gives an advantage to deal with the PAPR in a different direction.

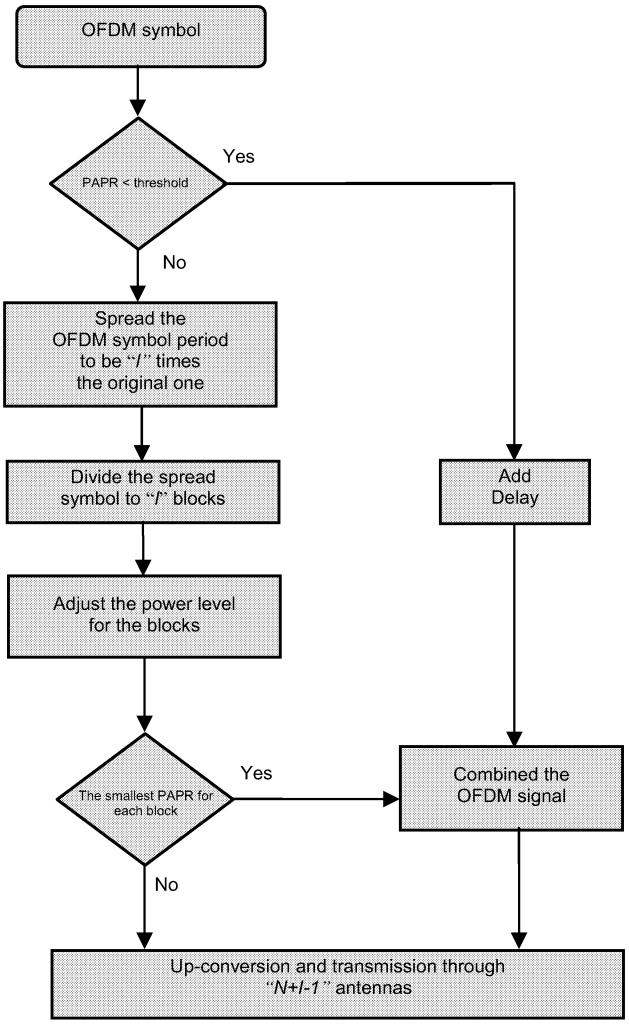


Figure 2: Eliminating the PAPR flowchart

Figure 2 shows the steps of this work. After the IFFT stage, the OFDM symbol is generated and it is possible to be affected by a high PAPR. Therefore, the data will go through a detector with a defined threshold to determine whether the symbol is affected or not. The affected symbol will go through a spreading coding technique after removing its guard interval. The idea from imposing a spreading coding technique comes from equation (8), since an increasing in the total number of subcarriers will increase the average power. From the definition of the PAPR a direct reduction can be achieved by increasing the average power of the signal. The next step is to transform the encoded symbol in to parallel blocks with the same

period as the conventional OFDM symbol. After that one of these blocks, which contains the lowest PAPR, will join the OFDM signal while the others go through the up-conversion block to be transmitted as extra information that can be used in the receiver to rebuild the original symbol.

However, to make the calculations simpler since most systems are implemented in digital format and the calculations in continuous is more complicated. Therefore, a discrete format from (6) must be obtained by sampling the modulated symbol at least two times the Nyquist rate. The new version of (6) is shown below:

$$\text{PAPR}= \frac{\max\{|f_v|^2\}}{\frac{1}{P}(\sum_{v=0}^{P-1}|f_v|^2)}, v = 0,1,\ldots, P \qquad (7)$$

Equation (7) can be written to be more general as shown in (8), especially when the sampling rate will be L times the Nyquist rate:

$$\text{PAPR}= \frac{\max\{|f'_v|^2\}}{\frac{1}{LP}(\sum_{v=0}^{P-1}|f'_v|^2)}, v = 0,1,\ldots,LP \qquad (7)$$

TURBO CODING

In the literature, there exist a number of methods for encoding, such as block coding, convolutional codes, parity check codes, etc. In this work, we consider the use of a turbo coding technique. In fact, the turbo coding technique was introduced in 1993 and many researchers in the wireless sector see turbo coding as the forward-error-correction (FEC) technology of choice for 3G applications. We note that many of the International Telecommunication Union (ITU) specifications recommend using turbo coding for 3G architectures (Schlegel C., and Perez L., 2003).

This type of coding is referred to the usage of recursive systematic convolutional (RSC) codes separated by an interleaver. In this turbo encoder architecture, the RSCs are concatenated in parallel fashion. The constituent codes are RSCs because they combine the properties of non-systematic codes and systematic codes. Furthermore, Generator matrices, which represent the RSCs, must be defined. A permuter can be included in the turbo encoder architecture to take each incoming block of N bits and rearrange them in some particular fashion depending on the permuter itself. For large block sizes, designers can use pseudo random interleavers, which can be considered as a permutation map, to rearrange bits. To create the pseudo random interleaver map: generate n random numbers and rearrange them in ascending order or descending order. The bit rate at the output of the turbo coding is depends on the coding rate, as an example if the coding rate is x_1 and input rate is y_1 bits/sec, then the output bit rate is (x_1*y_1) bits/sec.

NUMERICAL RESULTS AND DISCUSSION

A MATLAB simulation program was used to perform extensive simulations and to evaluate our work. A total of 256 bit per symbol was created for the purpose of simulation and it was run for 20 times to increase the performance. Tables 1-4 show the final simulation results each of which were obtained by taking the average of the 20 runs of the algorithm.

Table 1: The simulation results with BPSK modulation technique.

Mapping tech.	Coding rate	MIMO encoding	Guard Interval 0.25		
			Old PAPR (dB)	New PAPR (dB) Spreading rate 2	3
BPSK	1/2	STBC	6.9	0.24	0.108
		V-BLAST	8.9	0.145	0.102
	1/3	STBC	14.1	0.108	0.103
		V-BLAST	12.67	0.154	0.089

Table 2: The simulation results with QPSK modulation technique.

Mapping tech.	Coding rate	MIMO encoding	Guard Interval 0.25		
			Old PAPR (dB)	New PAPR (dB) Spreading rate 2	3
QPSK	1/2	STBC	11.4	0.142	0.091
		V-BLAST	9.337	0.131	0.121
	1/3	STBC	12.3	0.128	0.1276
		V-BLAST	11.08	0.125	0.124

Table 3: The simulation results with 16-QAM modulation technique.

Mapping tech.	Coding rate	MIMO encoding	Guard Interval 0.25		
			Old PAPR (dB)	New PAPR (dB) Spreading rate 2	3
16-QAM	1/2	STBC	13.3	0.147	0.132
		V-BLAST	12.42	0.129	0.126
	1/3	STBC	13.8	0.099	0.098
		V-BLAST	14.3	0.104	0.1

Table 4: The simulation results with 64-QAM modulation technique.

Mapping tech.	Coding rate	MIMO encoding	Guard Interval		
			0.25		
			Old PAPR (dB)	New PAPR (dB)	
				Spreading rate	
				2	3
64-QAM	1/2	STBC	14.24	0.118	0.1176
		V-BLAST	9.6	0.156	0.099
	1/3	STBC	9.51	0.147	0.131
		V-BLAST	11.67	0.154	0.138

The tables shown above summarise the achievement of a considerable PAPR reduction for different modulation and encoding techniques.

CONCLUSION

In this paper, we have presented our new technique to tackle the PAPR problem in the high data rate systems. Our approach depends on handling this problem by using a spreading technique. In our case, we use the Turbo coding to achieve this spreading. It is evident from the simulation results that our technique has significantly reduced the PAPR for the high data rate systems. This improvement in the systems performance have been maintained even when different combination of mapping and MIMO coding techniques were used. The results have also shown that there was slight improvement in the PAPR reduction when the spreading rate that is used in our technique has been increased.

Our technique shows far better results than the classical techniques which are not using the spreading method in tackling the PAPR problem.

REFERENCES

Cimini L. and Sollenberger N., "Peak-to-Average power Ratio reduction of an OFDM signal using partial transmit sequence", *IEEE Communications Letters,* vol. 4, no. 3, pp. 86-87, March 2000.

H. Ochiai and H. Imai ,"*On the distribution of the Peak-to-average power ratio in OFDM signals*", *IEEE Transaction on Communications*, pp. 282-289, (2001).

Nee R. and Prasad R., "OFDM wireless multimedia communications", Artech House Boston London, 2000.

Nee R. and Wild A., "Reducing the Peak-to-Average Power Ratio of OFDM", *IEEE Vehicular Technology Conference 1998.*

Schlegel C., and Perez L., "Trellis and Turbo coding", Wiley-IEEE Press, 2003.

Shepherd S., et al, "Asymptotic Limits in Peak Envelope Power Reduction by Redundant Coding in Orthogonal Frequency-Division multiplex Modulation", *IEEE Transaction on Communications*, pp. 5-10, (1998).

Stuber G., "Principles of mobile communications", second edition, Kluwer Academic Publishers, 2001.

Data Mining for Generating Predictive Models of Automatic Speech Recognition

Ayad T. Al-Zobaydi, Marwan M. Al-Akaidi and Robert I. John

Wireless & Multimedia Communications & Signal Processing Group

De Montfort University, Leicester, U.K.

KEYWORDS

Artificial Intelligence, Automatic Speech Recognition, Data Mining, Artificial Neural Networks, Hidden Markov Model

ABSTRACT

Automatic Speech Recognition (ASR) has been gaining wide interest in recent years. Many different methods has been defined for speech recognition, like Artificial Neural Networks (ANN), Hidden Markov Model (HMM), or even hybrid systems like ANN/HMM.

Speech Recognition is still a very cumbersome problem. There are quite a lot of improvements to be done in terms of accuracy, training and recognition time, as well as avoiding over training problem.

This paper presents a new method for speech recognition using Data Mining (DM) principles.

INTRODUCTION

Automatic Speech Recognition (ASR) is a technology that allows a computer to identify the words that a person speaks into a microphone or telephone. The "holy grail" of ASR research is to allow a computer to recognize with 100% accuracy all words that are intelligibly spoken by any person, independent of vocabulary size, noise, speaker characteristics and accent, or channel conditions [2].

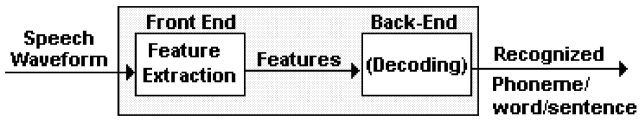


Figure (1): Phoneme identification stages

Contemporary Automatic Speech Recognition (ASR) systems are composed of a feature processing stage, which aims at extracting the linguistic message while suppressing non-linguistic sources of variability, and a classification stage (include language modelling), that identifies the feature vectors with linguistic classes, see Figure(1) [4].

The extraction level of current ASR systems converts the input speech signal into a series of low-dimensional vectors, each vector summarising the necessary temporal and spectral behaviour of a short segment of the acoustical speech input. The ultimate goal is to estimate sufficient statistics to discriminate among different phonetic units while minimizing the computational demands of the classifier. The lowest level, of most speech recognition systems, involves classifying short (e.g. 10 ms) windows, or frames, of speech waveforms to a given set of phoneme classes. The sequence of frames is then segmented and classified as a sequence of phonemes, which are further segmented and classified as a sequence of words [9].

In ASR, classification is done by "*deduction*" either logically or statistically. ASR also needs to have a "knowledge acquisition" capability (***by either abduction, or induction***). Thus, these requirements (deduction, abduction, and induction) make ***ASR potentially a problem that should be tackled Artificial Intelligence*** [1].

Because speech sound features (data) are random and continuous one can conclude that a probabilistic approach would help with classification in ASR. Also because of the sequential nature of the speech events, parallel classification is limited in speech continuous mode. Hidden Markov Model (HMM) is a statistically insensitive approach, and ANN is a knowledge insensitive approach. Therefore, among classification techniques HMM has proved itself to be a promising technique., The ANN approach has also shown some promising results [5]. A number of researchers try to solve the problem of each of HMM and ANN by proposing a hybrid system that involves use of each technique, known as ***HMM/ANN hybrid system, which seems the most suitable ASR architecture defined yet*** [1]. Although it appears to be promising, there are significant issues to do with accuracy, training time, and the over-training problem.

DATA MINING (DM)

Data mining refers to the mining or discovery of new information in terms of patterns or rules from vast amounts of data. It uses techniques from such area as machine learning, statistics, neural networks and genetic algorithms. Broadly speaking, the goals of data mining fall into the following classes: ***prediction, identification, classification, and optimization***. The knowledge discovered during data mining can be described in the ways: ***Association rules, Classification hierarchies (Regression, ANN, and Genetic Algorithms), Sequential patterns, Patterns within time series, Categorization and segmentation*** [10].

Data mining is defined also as a process that begins with data in some form and ends with knowledge in some form [11].

With DM tools, instead of assuming certain relationship between different variables in a data set (and studying these relationships one at a time), we can

determine the most significant factors that influence the outcome. DM does not need any hypotheses - it discovers hidden relationships and patterns automatically. DM has found numerous applications in banking, finance, marketing, telecommunications, and now in, for example, the semantic web [7].

A system is formally described by a model that may take the form of a mathematical expression or algorithm that provides a value based on input variables. Data mining models can be either descriptive or predictive. Data mining is conceived of as a process having two components: Discovery, during which meaningful patterns are detected in data and characterized formally (descriptive models); and exploitations, during which meaningful patterns are used to create useful applications (predictive models) [11].

Data mining models typically classify or predict, so are called predictive models. The term predictive does not suggest that there is any prognostication involved. Predictive models are "pattern classifiers". The process of building predictive models using data mining methods and tools is referred to as predictive modeling. Predictive models can be constructed in many ways, because there are many ways to approach the problem of exploiting patterns in data. Patterns can be processed using many techniques like polynomial regression, Knowledge Base Expert System (KBES), and ANN [11].

Data Mining Model for ASR (DMM4ASR)

A software system for speech recognition using DM methodology has been built to recognize phonemes. A sketch for DMM4ASR is illustrated in Figure (2). This system consists of four main modules: ***get-phoneme*** module, ***feature extraction*** module, ***Recognition*** module, and the ***Interface*** module. These modules have been implemented as functions using Matlab™ software. The data of the system is TIMIT database, which contains training and testing data. A brief description for each module of DMM4ASR system is now given.

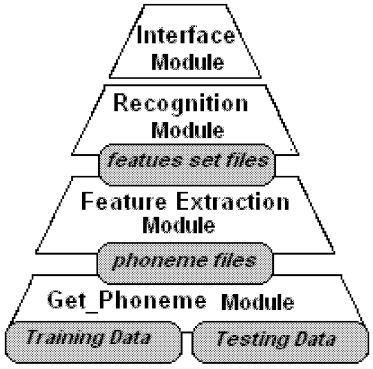


Figure (2): DMM4ASR components

- **Data**

This system uses the TIMIT acoustic-phonetic speech corpus. TIMIT was designed to provide speech data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of ASR systems. These data was prepared at the National Institute of Standards and Technology (NIST) with sponsorship from the Defense Advanced Research Projects Agency - Information Science and Technology Office (DARPA-ISTO) [13].

- **Get_Phoneme module**

The Hamming windowing technique is one of a number of tokenizing techniques of an utterance [5]. Hamming is used in this module for tokenizing an utterance of a statement to its phonemes, and stores each phoneme in a separate file. The input data to this module is a file from training or testing set of TIMIT, contains utterance of a statement, and the output of this module is a number of phoneme files (in text format).

- **Feature Extraction module**

There are many kinds of continuous data that provide a fertile field for data mining, such as audio, video and other types of signal data. In order for data mining to be performed on such data, "features" of the data must be extracted or determined through some mechanism [11].

This module implements Mel-scale Frequency Cepstral Coefficients (MFCC) technique and it aims to produce a feature set of the speech signal of a phoneme that will be used for recognition purpose. Figure (3) illustrate the flow diagram of MFCC computation [5][3].

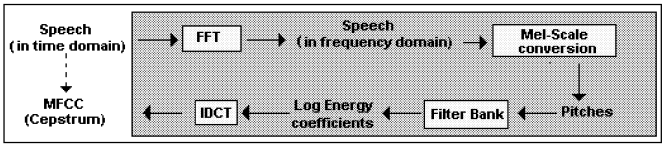


Figure (3): MFCC computation flow diagram

A function called "Get_MFCC[Cep]" has been built using Matlab.

- **Recognition module**

This module takes a phoneme features (stored in a file) and recognizes it. The recognition approach is of two stages, the first one is clustering (grouping), and the second stage is identification of the input features. The whole recognition approach uses multiple reference patterns instead of one reference pattern as in the usual methods.

The architecture of this model is hierarchical and consists of two levels; each one is a ANN which has a different function from the other. The super network is responsible for clustering the incoming features and decides which of the subnets would be used for the next stage of recognition. The second one is responsible for identification of an element inside its cluster. Figure (4) illustrate the architecture of the inter-network used in DMM4ASR.

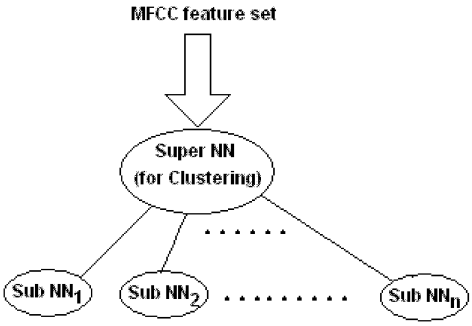


Figure (4): Architecture of hierarchical NN Recognition model

In ***supervised learning***, the NN learns to ***classify*** patterns according to instruction: a target pattern informs the network as to which class it should learn to classify an input pattern. With ***unsupervised learning*** there is no such instruction and the network is left to ***cluster*** (or group) patterns.

All of the patterns within a cluster will have something in common: they will be judged as being similar. The groups are referred to as clusters and the arrangement of clusters should reflect two properties:

- Patterns within a cluster should be similar in some way.
- Clusters that are similar in some way should be close together.

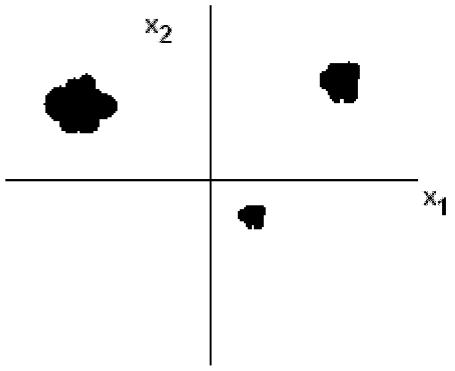


Figure (5): Data that form three clusters

Figure (5) shows a plot of two dimensional data that fall naturally into three clusters. In Figure (5), a pattern point can be said to fall naturally into a cluster if it is close to a point in the same cluster compared to a point belonging to another cluster. The Euclidean squared distance is typically used to measure how close (or similar) two points are, and is given by:

$$d_{pq} = \sum_{i}^{n} (x_{pi} - x_{qi})^2$$

where d_{pq} is the Euclidean squared distance between point p and point q, x_{pi} is the coordinate i for pattern p (similarly for q) and n is the number of dimensions. If there exists a vector, p_j for cluster j , located at the centeriod (the position that is the average of all patterns in a cluster) of each cluster in Figure (5) then a decision as to which cluster some vector x (seen as a point when plotted) belongs to is simply:

$$index\ (x) = min\ d(p_j, x), \quad for\ all\ j$$

Which returns the index of the cluster with the lowest Euclidean squared distance from vector x. The vectors p_j can be considered as prototypes for the clusters; these prototypes serve to represent the key features of a cluster.

A cluster algorithm is a statistical technique for discovering clusters in a set of data. There are many clustering algorithms of varying complexity. A simple approach to clustering is to assume that a certain number of clusters exist and randomly assign the coordinates for each prototype. Each vector in the data set is then assigned to its nearest prototype and the prototype updated (moved) to become the centeriod of all vectors assigned to that prototype [12].

The requirements of this model are:

- ***Data (speech features)***: two types of data, training data and testing data
- ***Learning algorithm***: defines the thresholds for clustering super-NN and classification for sub_NNs.
- ***Recognition algorithm***: of deductive approach, by clustering in first and identification later.

- **The Interface module**

The aim of this model is to provide the interface between the user and the DMM4ASR system. A graphical user interface has been used to make the system ease to use.

RESULTS

1. ***Designing:*** The main problem here is designing the topology of the ANNs.
2. ***Implementation***: This has been done using Matlab™
3. ***Accuracy***: This factor appears good with when the trained NN has been tested with number of test data belong to a number of different speakers. Table (1) shows the percentage of correct detection of some phonemes over our design.

		Training Data		
		/aa/	/iy/	/ah/
Testing Data	/aa/	**91.29%**	5.88%	2.83%
	/iy/	4.99%	**91.75%**	3.26%
	/ah/	2.1%	6.06%	**91.84%**

Table (1):Confusion matrix for testing datasets using 13 coefficients of 'MFCC'

4. ***Intermediate results***: these include the phoneme features, which have been calculated by feature extraction module. The phoneme signal has been tested by playing the segmented period from the whole utterance. Also reconstruction of the original FB output have been done for checking purposes. It has been done by multiplying the cepstral data by the transpose of the original Discrete Cosine Transform (DCT) matrix. This all works because we were careful to scale the DCT matrix so it was orthonormal.

5. ***Training***: although the training has been done in two stages it takes less time, and the over-training problem has been avoided by using less iterations in training.
6. ***Recognition Time***: due to the parallelism nature of the NN, time taken in the recognition is considered at least well.

CONCLUSION

1. This research introduces a ***new architecture of DM*** that consists of multiple ANNs of two stages hierarchical architecture.
2. ***The success of using DM methodology to construct ASR system***, due to the fact that ASR is one of AI systems, that requires both descriptive and predictive models which defined by DM techniques.
3. While HMM requires a manual definition for an automata of a series of phonemes (or words) as a prerequisite to build ASR system, our method does not since it is a kind of DM, hence ***it is more flexible than HMM***.
4. The ***promising advances showed by the results of running the DMM4ASR***, over easy of design, training & decoding time, and the accuracy of the output.

REFERENCES

1. Ayad T. Al-Zobaydi and Marwan M. Al-Akaidi, “Speech Recognition by Computer-A Survey Article", MESM04.
2. Daniel Jurafsky & James H. Martin, "Speech and Language Processing", Prentice-Hall, Inc., 2000.
3. S.Fekkai, M. Al-Akaidi, "Fractal based speech recognition and synthesis", Ph.D. study, De Montfort University/ Leicester/U.K., 2002.
4. Joseph Picone, “Fundamentals of speech recognition”; lecturers on web site of URL: www.isip.msstate.edu/resources/cources/ece_8463.
5. Marwan M. Al-Akaidi; "Fractal Speech processing"; ISBN: 0521814588; Cambridge University Press; 2004
6. Melvyn J. Hunt; “signal representation”; web-page lecture titled, part of "Survey of the state of the Art in Human Language Technology"; published by Centre for Spoken Language Understanding (CSLU) - Oregon Graduate Institute of Science and Technology; URL:"http://www.cslu.ogi.edu/HLTsurvey"; 1996.
7. Michael Negnevitsky; "Artificial Intelligence - A guide to intelligent systems": second edition; ISBN: 0321204662; Addison Wesely; 2005.
8. Olivia Parr Rud; "Data Mining Cookbook"; ISBN: 0-471-38564-6; Wiley publishing company 2001.
9. Rabiner L. R. And Juang B.H., "Fundamentals of Speech Recognition", PTR Prentice Hall, 1993.
10. Ramez Elmasri and Shamkant B. Navathe, "Fundamentals of Database Systems", third edition, ISBN: 0-8053-1755-4, Addison Wesley 2000.
11. Rhonda Delmater & Monte Hancock; "Data Mining Explained"; ISBN: 1-55558-231-1; Butter-Heinemann (member of the Reed Elsevier group) 2001.
12. Robert Callan; "The Essence of Neural Networks"; ISBN: 0-13-908732-X; Prentice Hall 1999.
13. Training and Test Data and Speech Header Software NIST Speech Disc CD1-1.1; The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus;1990.

SIMULATION AND FPGA IMPLEMENTATION OF A SIMPLE COMPUTER

Dr. Kasim M. Al-Aubidy, Ra'ed F. Al-Bader, Ala'a A. Smadi
Computer Eng. Dept,
Philadelphia University,P O Box 1,
Jordan, 19392
E-mail: kma@philadelphia.edu.jo

KEYWORDS
Computer systems, Processor design, Simulators, Microprogramming.

ABSTRACT

FPGA technology offers the potential of designing high performance systems at low cost. FPGAs have been used for many computational tasks, and this paper presents the micro-operation simulation of a basic computer and its implementation on an FPGA. Also, it covers the design of an assembler for the designed computer, which can be used for educational purposes. Such implementation offers good code density, easy customization, easily developed software, small area, and high performance at low cost.

INTRODUCTION

Field programmable gate arrays (FPGAs) consist of programmable logic blocks, which can each implement a small amount of digital logic, and programmable routing which allows the logic block inputs and outputs to be connected to form larger circuits (Ciletti 2002). FPGAs have become a popular technology for creating digital systems since they can lead to a shorter time-to-market for designs than application-specific circuits (ASICs) and allow design modifications to be made after system creation (Welch and Carletta 2000, Krid and Masmoudi 2005). The primary method used for validating a design with most FPGA design flows is simulation.
A Hardware Description Language (HDL) model is more precise and concise than a logic circuit model, for system design and realization. It permits the user to create a design without having to first choose a device for implementation (Muresan 1997, Mano 1993). An HDL model is the key for the transition from System-On-a-Board to System-On-a-Chip.
This paper presents the simulation and FPGA realization of a basic computer and its assembler which can be used for educational purposes. This simulation is a set of micro-operations that represent the register transfer statements of all operations that can be implemented.

BASIC COMPUTER DESIGN AND SIMULATION

A basic computer is designed with minimum hardware components to perform the basic tasks that a computer should perform (Mano 1993). The organization of the computer is defined by its internal registers.

The internal organization of a digital system is defined by sequence of micro-operations that perform on data stored in its registers. The computer is capable of executing various micro-operations and can be instructed to perform a sequence of operations.
Instruction codes together with data are stored in memory. The computer reads each instruction from memory and places it in a control register. The control unit then interprets the binary code of the instruction and proceeds to execute it by issuing a sequence of micro-operations.
An instruction code is a group of bits that instruct the computer to perform a specific operation. It is made up of 16 bits, and divided into three parts:

- Two bits (I1,I0) to specify the addressing mode.
- Four bits binary code to specify the operation.

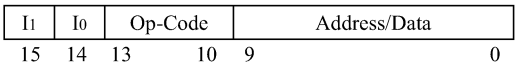


Figure 1: Instruction format.

Table 1 illustrates the addressing mode of the proposed computer, which is used to:

- Reduce the number of bits in the address field of the instruction, and
- Give user flexibility in dealing with counters, pointers …etc

Table 1: Addressing Modes.

I_1	I_0	Addressing mode
0	0	Immediate addressing *($ address)*
0	1	Direct addressing
1	1	Indirect addressing *(# address)*

The computer needs registers for manipulating data and a register for holding a memory address. Nine registers are required for the proposed computer, as shown in Table 2. Some registers (such as AC, MAR, and MBR) may receive data from several multiplexed sources.

Table 2: List of Registers for the Basic Computer.

Register	Bits No.	Function
Memory Buffer Reg. (MBR)	16	Holds memory word
Memory Address Reg. (MAR)	10	Holds address for memory
Address Reg. (AR)	10	Holds operand/instruction add.
Accumulator (AC)	8	Processor register
Counter Reg. (CR)	8	Holds count for loops
Program Counter (PC)	10	Holds address of instruction
Operation Reg. (OPR)	4	Holds code of operation
Input Reg. (INP)	8	Holds input character
Output Reg. (OUTR)	8	Holds output character

Table 3: Basic computer instruction formats

MEMORY REFERENCE INSTRUCTIONS				
Symbol	Code for the 6 left most bits			Description
	I1I0=00	I1I0=01	I1I0=11	
NOP	000000	010000	110000	No operation
ADD	000001	010001	110001	Add memory word to AC
SUB	000010	010010	110010	Subtract memory word from AC
AND	000011	010011	110011	AND memory word to AC
OR	000100	010100	110100	OR memory word to AC
XOR	000101	010101	110101	XOR memory word to AC
LDA	000110	010110	110110	Load memory word to AC
STA	000111	010111	110111	Store content to AC in memory
BUN	001000	011000	111000	Branch unconditionally
BSA	001001	011001	111001	Branch and save return address
DSZ	001010	011010	111010	Decrement and skip if zero
LDC	001011	011011	111011	Load CR
RET	001100	011100	111100	Return
BZ	001101	011101	111101	Branch if zero
BC	001110	011110	111110	Branch if carry
REGISTER REFERENCE INSTRUCTIONS (Hexadecimal code)				
CLA	3C01			Clear AC
CLS	3C02			Clear all status flags
CMA	3C04			Complement AC
SRA	3C08			Shift right AC
SLA	3C10			Shift left AC
INC	3C20			Increment AC
HALT	3C40			Terminate program
INPUT-OUTPUT REFERENCE INSTRUCTIONS				
INP	BC01			Input character to AC
OUT	BC02			Output character from AC
SKI	BC04			Skip on input flag
SKO	BC08			Skip on output flag
ION	BC10			Interrupt on
IOF	BC20			Interrupt off
SFI	BC40			Set input flag
SFO	BC80			Set output flag

The basic computer has eight registers, a memory unit, and a control unit. Paths must be provided to transfer information from one register to another and between memory to registers. A more efficient scheme for transferring information in a system with many registers is to use a common bus. A multiplexer can be used to design the common bus. The connection of the registers to the common bus system is shown in Fig. 2.

Computer Instructions:

The basic computer has three basic instruction code formats. A memory-reference instruction uses ten bits to specify either an address or an operand and two bits to specify the addressing mode (I_0,I_1). For immediate addressing it is 00, 01 for direct addressing, and 11 for indirect address. The register-reference instructions are recognized by the operation code 1111 and a 00 in the left most bits of the instruction. A register-reference instruction specifies an operation on the AC register. An operand from memory is not needed; therefore, the other 10 bits are used to specify the operation to be executed. Similarly, an input-output instruction does not need a reference to memory and is recognized by the operation code 1111 and 11 in the left most bits of the instruction. The remaining 10 bits are used to specify the type of the input-output operation. This technique allows having up to 35 different operations, as given in Table 3.

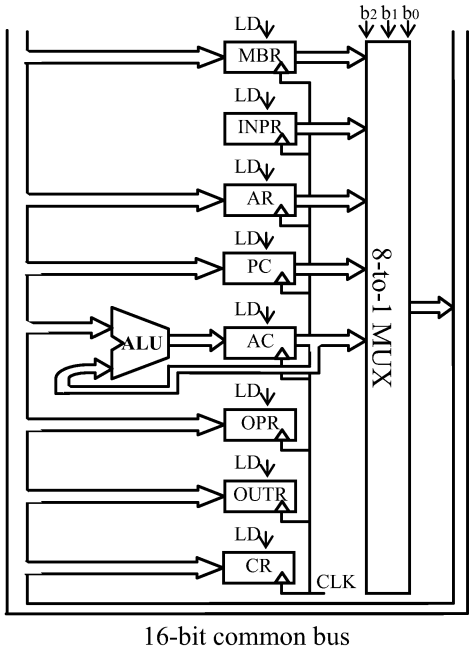


Figure 2: Basic computer registers connected to a common bus.

Timing and Control:

The timing for all registers in the basic computer is controlled by a master clock generator. The control signals are generated by the control unit and provide control inputs for the multiplexer in the common bus, control inputs in processor registers, and micro-operations for the accumulator. The control unit, Fig. 3, consists of three decoders, a sequence counter, and a number of control logic gates. The control unit gets the operation code from the OPR through a 4x16 decoder. Bits 14 & 15 of the instruction code are transferred to two flip-flops designated by the symbols I_1 & I_0. Bits 0 through 9 are applied to the control logic gated directly from the MBR through the common bus. The outputs of the counter are decoded into 4 timing signals T0 through T3. The sequence counter (*SC*) can be incremented or cleared. Most of the time, the counter is incremented to provide the sequence of timing signals out of the 2x4 decoder. Once, the counter is cleared, causing the next active timing signal to T0. The decoder is used to determine the cycle to be performed.

INSTRUCTION CYCLES

A program residing in the memory unit of the computer consists of a sequence of instructions. Each instruction cycle is subdivided into a sequence of sub-cycles. The value of three flip-flops is entered into a decoder to determine the cycle to be served, as illustrated in Table 4. In this computer each instruction cycle is divided into the following five sub-cycles:

Table 4: Cycles Combination.

Flip-flops			
G	**F**	**R**	**Cycle**
0	0	0	Fetch & Decode
0	0	1	Indirect
0	1	0	Direct
0	1	1	Execution
1	0	0	Interrupt

Fetch and Decode Cycle

Initially, the program counter is loaded with the address of the first instruction in the program. The sequence counter is cleared to 0, providing a decoded timing signal T0. After each clock pulse, sequence counter is incremented by one, so that the timing signals go through a sequence T0, T1, T2 and T3. The micro-operations for the fetch and decode cycle can be specified by the following statements:

$$
\begin{aligned}
&C_0T_0 : PC \rightarrow MAR \\
&C_0T_1 : M(MAR) \rightarrow MBR,\ PC+1 \rightarrow PC \\
&C_0T_2 : MBR < I_0, I_1 > \rightarrow (I_0, I_1), \\
&\qquad MBR < OP_Code > \rightarrow OPR, \\
&\qquad MBR < Address > \rightarrow AR \\
&\overline{I_0}C_0T_3 : \quad 1 \rightarrow R,\ 1 \rightarrow F, 0 \rightarrow G \qquad (Execute Cycle) \\
&\overline{I_1}I_0C_0T_3 : O \rightarrow R,\ 1 \rightarrow F, 0 \rightarrow G \qquad (Direct Cycle) \\
&I_1I_0C_0T_3 : 1 \rightarrow R,\ 0 \rightarrow F, 0 \rightarrow G \qquad (Indirect Cycle)
\end{aligned}
$$

Indirect Cycle

At this cycle, the effective address of the operand is to be read from memory. The AR holds the address of memory word which contains the effective address of the operand.

$$
\begin{aligned}
&C_1T_0 : AR \rightarrow MAR \\
&C_1T_1 : M(MAR) \rightarrow MBR \\
&C_1T_2 : MBR < Address > \rightarrow AR \\
&C_1T_3 : 0 \rightarrow R,\ 1 \rightarrow F, 0 \rightarrow G \qquad (Direct Cycle)
\end{aligned}
$$

Direct Cycle

The effective address of the operand may be read during two time pulses. Therefore, to disable the delay of waiting for T3, the sequence counter may be cleared at time T2. Thus, the next time pulse will be T0 of the execution cycle and not T3 of the indirect cycle.

$$
\begin{aligned}
&C_2T_0 : AR \rightarrow MAR \\
&C_2T_1 : M(MAR) \rightarrow MBR, \\
&C_2T_2 : 1 \rightarrow R,\ 1 \rightarrow F,\ 0 \rightarrow G, \\
&\qquad 0 \rightarrow CC \quad (Execution \quad Cycle)
\end{aligned}
$$

Execute cycle

At this cycle, the fetched instruction (register-reference, memory reference, or input-output-reference) is executed. The type of the instruction is decided according to the following table:

Table 5: Combinations of Instruction Types

q15	I0	I1	Instruction Type
0	X	X	MRI
1	0	0	RRI
1	0	1	IOI

Interrupt Cycle

The interrupt cycle is initiated after the execute cycle if the interrupt flip-flop (*INF*) is equal to 1. The flip-flop is set to 1 manually using a switch, and it may be set during the execution of the program by the instruction ION. The flip-flop is reset to 0 whenever the interrupt is served or by the instruction *IOF*.

This basic computer serves the interrupt by saving the next sequential instruction in memory address 0, and then it starts execution from address 1 in the memory. The micro-operations required for this instruction are:

$$
\begin{aligned}
&C_4T_0 : PC \rightarrow MBR < Data >, 0 \rightarrow INF \\
&C_4T_1 : 0 \rightarrow PC, 0 \rightarrow MAR \\
&C_4T_2 : MBR < Data > \rightarrow M(MAR), PC+1 \rightarrow PC \\
&C_4T_3 : 0 \rightarrow R, 0 \rightarrow F, 0 \rightarrow G \qquad (Fetch Cycle)
\end{aligned}
$$

Table 6: Micro-operations of all instructions.

Instr.	Micro-operations.
Memory Reference Instructions (MRIs)	
NOP	No micro-operation
ADD	$q_0C_3T_0 : AC + MBR < Data > \rightarrow AC$
SUB	$q_1C_3T_0 : AC - MBR < Data > \rightarrow AC$
AND	$q_2C_3T_0 : AC \cap MBR < Data > \rightarrow AC$
OR	$q_3C_3T_0 : AC \cup MBR < Data > \rightarrow AC$
XOR	$q_4C_3T_0 : AC \oplus MBR < Data > \rightarrow AC$
LDA	$q_5C_3T_0 : MBR < Data > \rightarrow AC$
STA	$q_6C_3T_0 : AR \rightarrow MAR, AC \rightarrow MBR < Data >$ $q_6C_3T_1 : MBR < Data > \rightarrow M(MAR)$
BUN	$q_7C_3T_0 : AR \rightarrow PC$
BSA	$q_8C_3T_0 : PC \rightarrow MBR < Address>, AR \rightarrow MAR$ $q_8C_3T_1 : MBR < Address> \rightarrow M(MAR),$ $AR + 1 \rightarrow AR$ $q_8C_3T_2 : AR \rightarrow PC$
DSZ	$q_9C_3T_0 : MBR < Data > -1 \rightarrow MBR < Data >$ $q_9C_3T_1 : MBR < Data > \rightarrow M(MAR),$ $MBR < Data > \rightarrow AC$ $q_9C_3T_2 : IF(AC = 0)\ Then\ PC + 1 \rightarrow PC$
LDC	$q_{10}C_3T_0 : MBR < Data > \rightarrow CR$
BZ	$q_{11}C_3T_0 :$ if $ZF = 1$ then $PC + 1 \rightarrow PC$ else $MBR < Add > \rightarrow PC$
BC	$q_{12}C_3T_0 :$ if $CF = 1$ then $PC + 1 \rightarrow PC$ else $MBR < Add > \rightarrow PC$
Register Reference Instructions (RRIs)	
CLA	$rB_0T_0 : 0 \rightarrow AC$
CLS	$rB_1T_0 : 0 \rightarrow$ All Status Flags
CMA	$rB_2T_0 : \overline{AC} \rightarrow AC$
SRA	$rB_3T_0 : SHR(AC) \rightarrow AC, 0 \rightarrow AC(7)$
SLA	$rB_4T_0 : SHL(AC) \rightarrow AC, 0 \rightarrow AC(0)$
INC	$rB_5T_0 : AC + 1 \rightarrow AC$
HALT	$rB_6T_0 : 0 \rightarrow$ StartIndex
Input/Output Reference Instructions (IORIs)	
INP	$pB_0T_0 : INPR \rightarrow AC, 0 \rightarrow FGI$
OUT	$pB_1T_0 : AC \rightarrow OUTR, 0 \rightarrow FGO$
SKI	$pB_2T_0 : IF(FGI = 1)$ then $PC + 1 \rightarrow PC$
SKO	$pB_3T_0 : IF(FGO = 1)$ then $PC + 1 \rightarrow PC$
ION	$pB_4T_0 : 1 \rightarrow INF$
IOF	$pB_5T_0 : 0 \rightarrow INF$
SFI	$pB_6T_0 : 1 \rightarrow FGI$
SFO	$pB_7T_0 : 1 \rightarrow FGO$

Note that:

- $q_{15}\overline{I_1}\,\overline{I_0}C_3 = r$ (Common to all RRIs).
- $MBR(n) = B_n$ [MBR (0-9) that specifies the operation].
- $q_{15}I_1\overline{I_0}C_3 = p$ (Common to all input-output instructions).
- $MBR(n) = B_n$ [MBR (0-9) that specifies the operation].

Complete Computer Description

The instruction cycle flowchart is shown in Fig. 4. Figure 5 shows the schematic diagram of the basic computer. It is compound of one 16-bit register, three 10-bit registers, four 8-bit registers, one 4-bit register, eight D-Flip-flops, one 1024x16-bit RAM, one ALU, one 8-bit inverter, two 2-to-1 Multiplexer, one 4-to-1 Multiplexer, one 4-to-16 decoder, one 3-to-8 decoder, one 2-to-4 decoder, and one 2-bit sequence counter.

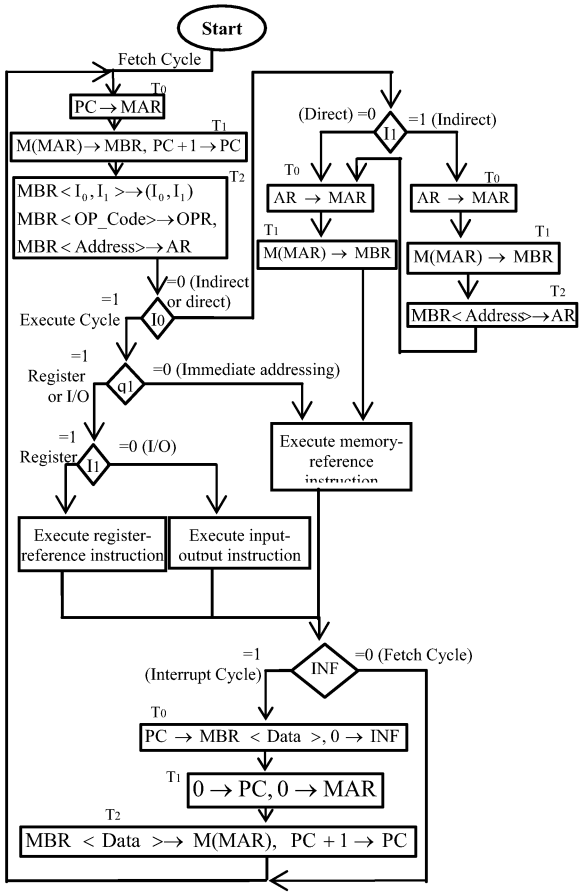


Figure 4: Flowchart of instruction cycle.

ASSEMBLER DESIGN

A two pass assembler has been designed and implemented to write assembly programs and use the output of the assembler to run these programs on the basic computer.

Input and Output Files

Figure 6 shows the files used as input and those generated as output by the assembler. These are;

- Source File (input): It is a text file containing the source program to be assembled. It has a *".asm"* extension. It consists of two segments, the code segment followed by the data segment. The data segment starts with "data:".
- Binary Code File (output): It contains the assembled statements represented in binary form. This file is stored in the block memory, and it has a *".dat"* extension.

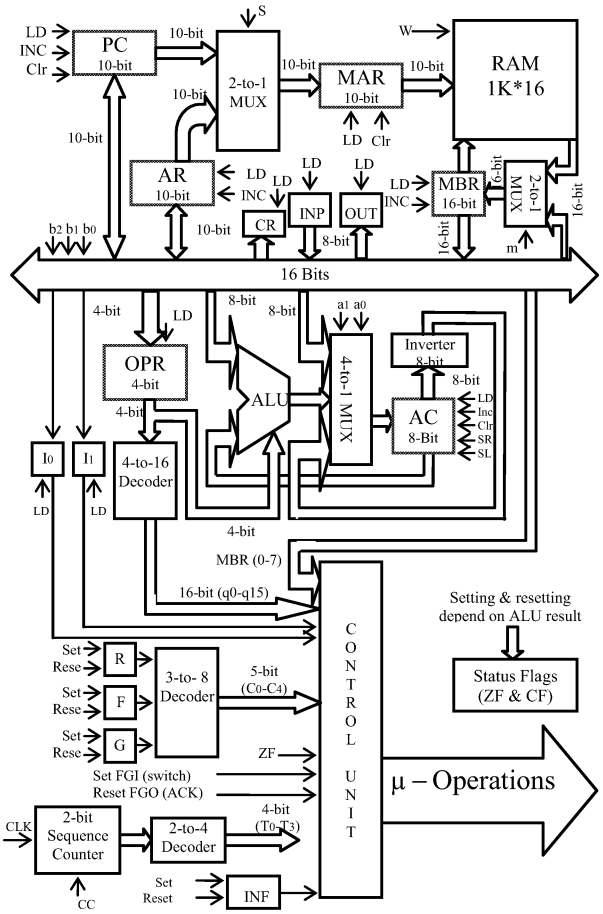


Figure 5: Basic computer block diagram.

- Hex Code File (output): It contains the assembled statements in hexadecimal form. This file is stored in the external memory, and it has a *".mem"* extension.
- Listing file (output): It consists of the source file statements, the assembled code, and the Branch Vector Table *BVT*. This file has a *".lst"* extension.

The Basic Computer Assembly Language character set consists of the following subset of the standard ASCII character set:

- Lower-case letters (a to z).
- Digits 0 through 9.
- Blanks (ASCII 32).

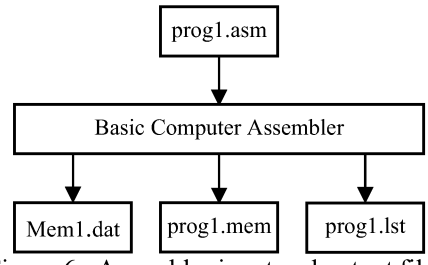


Figure 6: Assembler input and output files.

Assembler Instructions

The Basic Computer Assembly Language consists of lines of text where each line contains only one instruction and an optional comment. An instruction takes one or zero operands. Table 7 shows all instructions and their appropriate operands and addressing modes.

Table 7: Possible Types for Each Assembly Instructions.

Instr.	Type of operand	Function
nop	-	No operation
add	Immediate data/Add.	Add operand to AC
sub	Immediate data/Add.	Subtract operand from AC
and	Immediate data/Add.	AND operand to AC
or	Immediate data/Add.	OR operand to AC
xor	Immediate data/Add.	XOR operand to AC
lda	Immediate data/Add.	Load operand to AC
sta	Address	Store AC content in memory
bun	Lable	Branch unconditionally
bsa	Lable	Branch and save return add.
dsz	-	Decrement and skip if zero
ldc	Immediate data/Add.	Load CR
ret	-	Return
bz	Lable	Branch if zero
bc	Lable	Branch if carry
cla	-	Clear AC
cls	-	Clear all status flags
cma	-	Complement AC
sra	Immediate data	Shift right AC
sla	Immediate data	Shift left AC
inc	-	Increment AC
halt	-	Terminate program
inp	-	Input character to AC
out	-	Output character from AC
ski	-	Skip on input flag
sko	-	Skip on output flag
ion	-	Interrupt on
iof	-	Interrupt off
sfi	-	Set input flag
sfo	-	Set output flag

To branch to a line of assembly code, a symbolic destination must be placed at the beginning of the line of code. The comment symbol is //. A line that begins with // is considered a comment line. It is printed to the list (.lst) file but will not be encoded into the hex (.mem) and the binary (.dat) files. Comments may be added to lines that contain program code.

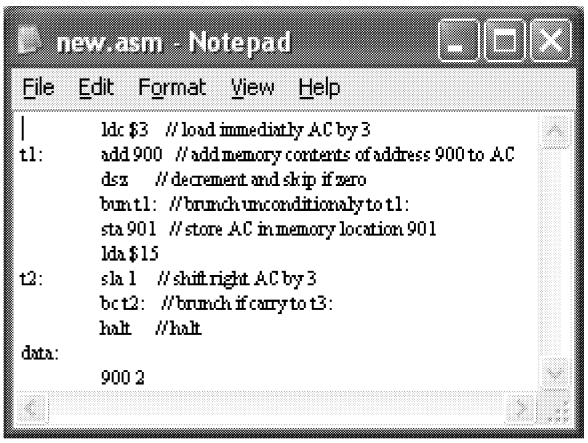


Figure 7: Example of a source code.

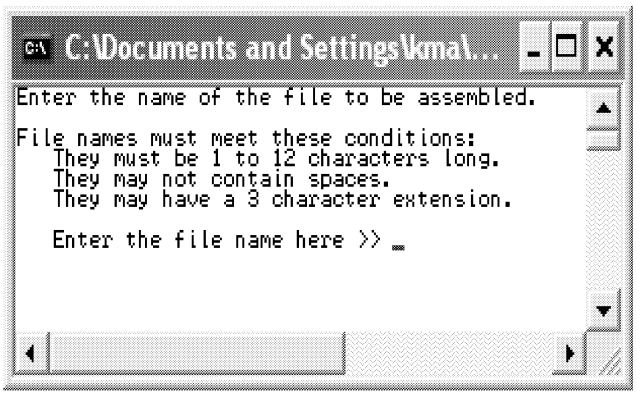


Figure 8: Assembler Window.

Example:

To use the assembler, follow these steps:

1. Write the program in Basic Computer Assembly language using a text editor such as Microsoft Notebook, see Fig. 7.
2. Save the file as text only, using a "*.asm*" extension.
3. Make sure your assembly language file is in the same directory as the Basic Computer Assembler.
4. Using Windows Explorer or a command window, start the assembler. Figure 8 shows the assembler window.
5. When the assembler comes up, enter the name of your file, being sure to include the *.asm* extension.

Then three new files are generated in the directory. The hex and list files will have the same name as the assembly language file with extensions *.mem* (see Fig. 9), and *.lst* (see Fig. 10). The binary file have a "*mem1.dat*" name, as in Fig. 11.

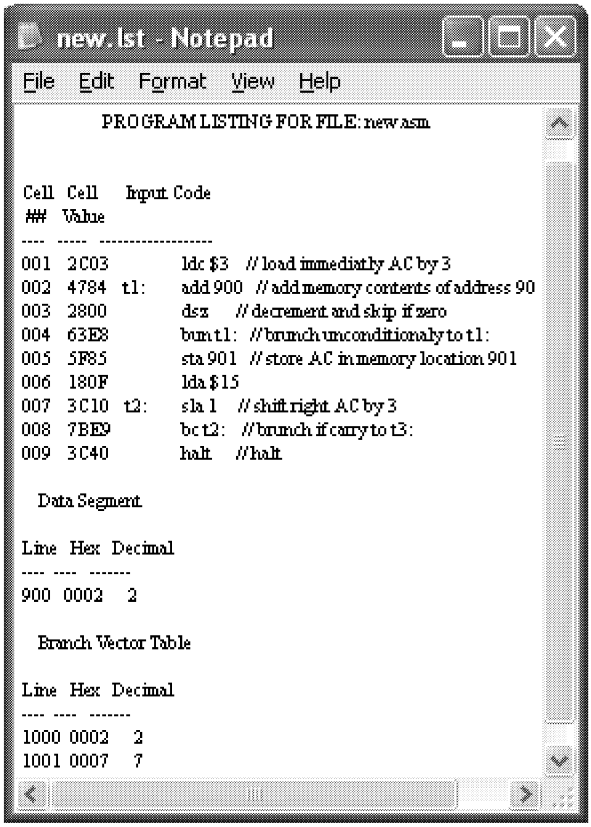


Figure 9: Listing file.

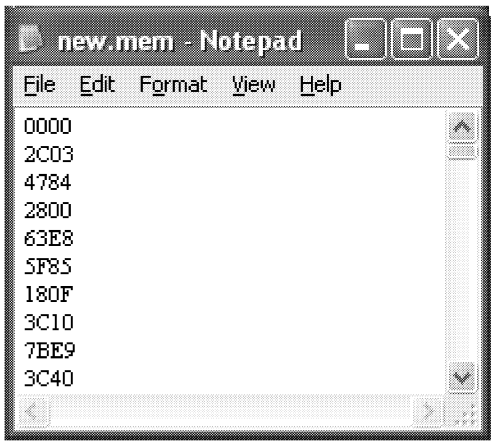


Figure 10: Beginning of Hex file.

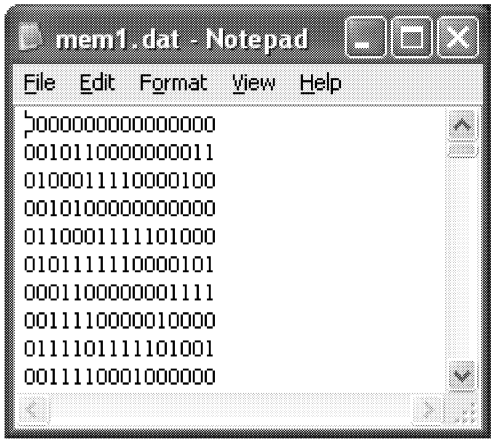


Figure 11: Beginning of mem1.dat file.

FPGA IMPLEMENTATION

Field programmable gate arrays (FPGAs) are a class of programmable logic devices based on an array of logic cells surrounded by a periphery of input/output cells. These programmable integrated circuits can be programmed in the field to implement specific design function.

The basic computer architecture is created from the ground up as a scalable architecture, covering the basic operations in 16-bit processor domain. Figure 12 illustrates the FPGA board (Xilinx Spartan-3).

The basic computer uses a four-character, seven segment LED display controlled by FPGA user-I/O pins. The interface for the seven-segment display is relatively simple, so it is used to display contents of system registers.

Four push button switches are used to moving-up registers sequence up/down, system reset, and user clock.

There are eight slide switches (SW_0-SW_7) in the system. Switches (SW_{4-7}) are used as input data, switches (SW_1 & SW_2) are used as external interrupts, while switch (SW_0) is used to indicate if the clock is a system clock or a user clock. Figure 13 shows the design procedure of the basic computer.

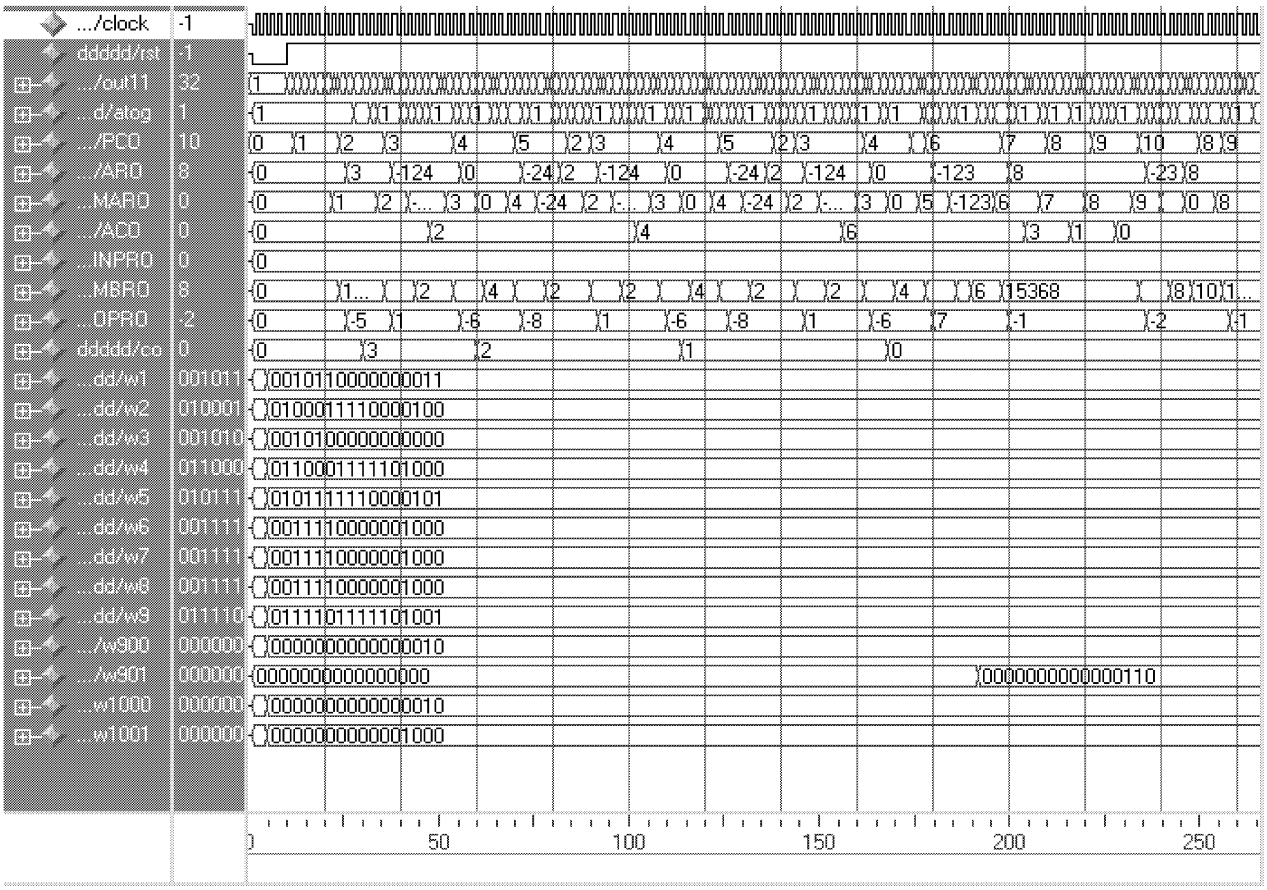


Figure 12: Waveforms Generated by the Program in Fig. 7.

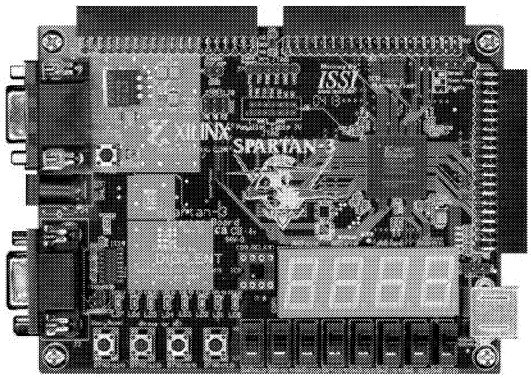


Figure. 13: Spartan-3 board.

Table 8 shows the different characteristics of the FPGA board we have implemented on.

Table 8: Device Characteristics

Device: Xilinx Spartan-3 XC3S200FT256 Slices: 3584, 256-ball thin Ball Grid Array				
System gates	Logic cells	Multipliers blocks (18*18)	Digital Clock Managers	Max I/O Signals
200 k	4320	12	4	173
Select RAM		In-system programmable configuration PROM	Fast Asynchronous SRAM	
18K-bits Blocks	Max RAM K-bits			
12	216	2 M-bits	1 M-byte	

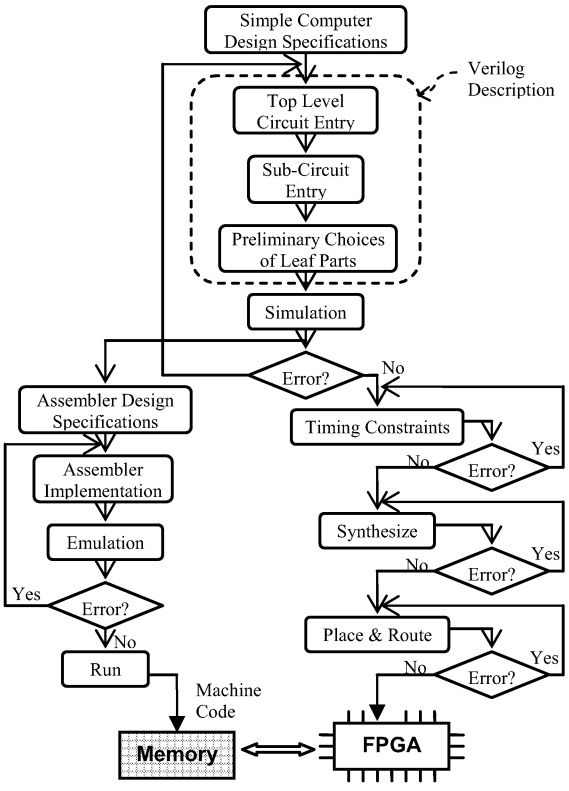


Figure 14: Design Flow Map.

Table 9: Macro Statistics

ROMs	16x7-bit
Registers	17
Counters	3
Multiplexers	7
Tri-states	27
Decoders	2
Adders/Subtractors	4

Tables 9 and 10 show the macro statistics and FPGA resources used respectively.

Table 10: Utilization Summary

Item	Used	Available	Percentage
Slices	433	3584	12%
Slice Flip Flops	114	7168	1%
4 input LUTs	802	7168	11%
Bonded IOBs	23	173	13%
GCLKs	2	8	25%

The circuit delay implemented on the FPGA is mostly due to routing delays, rather than logic block delays (Betz 2005). The average connection delay for this design is: 6.357 ns, the maximum pin delay is: 12.054 ns, the maximum operating frequency is 73.465 MHz, and the total memory usage is 90032 Kbytes.

CONCLUSIONS

- In this paper, the simulation of a basic computer and the implementation of an assembler has been presented.
- The micro-operation of the computer design and assembler design are implemented on an FPGA, since it offers good code density, easy customization, easily developed software, high performance and small area.
- The codes of the various modules are implemented and tested. The complete system can be achieved by linking up the modules together.
- The system has been tested with a program which utilizes every instruction as well as exercises the critical paths of the chip. It was found to be operational up to 73.465 MHz.

REFERENCES

Betz V. & J. Rose 2005, "FPGA Routing Architecture: Segmentation & Buffering to Optimize Speed & Density", *http://www.eecg.toronto.edu/~jayar/pubs/betz/fpga99 betz.pdf*

Ciletti M. D. 2002, "Advanced Digital Design with the Verilog HDL", *Prentice Hall,.1st edition, USA.*

Digilentic Co. 2004, "Spartan-3 Starter Kit Board User Guide", *April 26, http://www.digilentinc.com.*

Krid M. & D. S. M. Masmoudi 2005, "FPGA Implementation of a Feedforward Neural Network", *3rd Intr. Conf. on Systems, Signals & Devices, Vol.4, Tunisia.*

Mano M. M. 1993, "Computer System Architecture", *Prentice-Hall. 3rd edition, USA.*

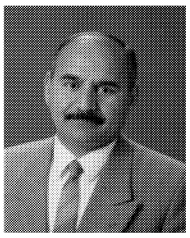
Muresan V, D. Crisu, & X. Wang 1997, "From VHDL to FPGA: A Case Study of a Fuzzy Logic Controller" *Proceedings of the Intr. Conf. of Young Lectures & PhD Students, pp(83-90), , Hungary.*

ST Microelectronics 2004, "Design Guid for Xilinx FPGA Power Management Systems", *Version 1.2, pp(1-11).*

Welch J.T. & J. Carletta 2000, "A Direct Mapping FPGA Architecture for Industrial Process Applications", *Proceedings of the 2000 IEEE Intr. Conf. on Computer Design: VLSI in Computers & Processors, ICCD2000, USA.*

.

AUTHOR BIOGRAPHY



Kasim Al-Aubidy received his BSc and MSc degrees in control and computer engineering from the University of Technology, Iraq in 1979 and 1982, respectively, and the PhD degree in real-time computing from the University of Liverpool, England in 1989. He is currently an associative professor in the Department of Computer Engineering at Philadelphia University, Jordan. His research interests include fuzzy logic, neural networks, genetic algorithm and their real-time applications. He is the winner of Philadelphia Award for the best researcher in 2000. He is also the chief editor of the Asian Journal of Information Technology. He has coauthored 3 books and published 54 papers on topics related to computer applications.

AUTHOR LISTING

AUTHOR LISTING

Aboud A. **64**
Acaccia G.M..... **35/41**
Agha A. **27**
Al-Akaidi M..... **101/143/147**
Al-Aubidy K.M. **151**
Al-Bader R.F. **151**
Alchaita M. **101**
Al-Khatib N..... **106**
Al-Zobaydi A.T. **147**
Amer G.M. **111**
Atri M. **126**

Ben Aouicha I..... **134**
Ben Halima D..... **49**
Ben Halima R..... **12**
Ben Salem F. **79**
Bodemann C.D. **5**

Chibani R. **95**
Chtourou M. **49/90/95**

Daoud O. **143**
Derbel H.B.J..... **85**
Derbel N..... **20/64**
Dia D..... **126**
Djemel M..... **70**
Drira K..... **12**

Ghommen J. **20**

Hamza R..... **64**

Irvine M..... **5**
Ivins J. **101/143**

Jmaiel M. **12**
John R.I. **147**

Kallel A. **119**
Khalkhali A..... **59**
Khanfir I. **119**

Masmoudi A..... **79**
Masmoudi D.S. **134**
Masmoudi N.K. **70**
Mechichi K. **64**
Mehran M. **59**
Michelini R.C. **35/41**
Mnif F..... **20**

Nariman-zadeh M.A..... **59**

Qualich N..... **35/41**

Rahhal J.S. **106**
Rekik C. **70**

Smadi A. **151**

Tawalbeh N. **27**
Tounsi-Rekik L..... **90**
Tourki R. **126**
Tray T. **134**